PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/11, 15/12, 15/62, C07K 14/505, 16/22, G01N 33/50, 33/53, A61K 48/00, 38/18, A61P 7/06

A1

(11) International Publication Number: WO 00/68376

(43) International Publication Date: 16 November 2000 (16.11.00)

(21) International Application Number: PCT/US00/12370

(22) International Filing Date: 5 May 2000 (05.05.00)

(30) Priority Data:

09/307,307	7 May 1999 (07.05.99)	US
Not furnished	28 March 2000 (28.03.00)	US
Not furnished	19 April 2000 (19.04.00)	US

(71) Applicant *(for all designated States except US)*: GENENTECH, INC. [US/US]; 1 DNA Way, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: DESAUVAGE, Frederic [BE/US]; 187 Shooting Star Isle, Foster City, CA 94404 (US). HENNER, Dennis, J. [US/US]; 297 Angelita Avenue, Pacifica, CA 94044 (US).

(74) Agent: ALTMAN, Daniel, E.; Knobbe, Martens, Olson & Bear, LLP, 16th floor, 620 Newport Center Drive, Newport Beach, CA 92660–8016 (US).

(81) Designated States: AE, AG, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, DZ, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KR (Utility model), KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NOVEL CHIMPANZEE ERYTHROPOIETIN (CHEPO) POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

```
human   MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE
        **************************************************
chepo   MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE

human   NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA
        ********************************** ****************
chepo   NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVRQQAVEVWQGLALLSEA

human   VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD
        ****************************************** ********
chepo   VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAXKEAISPPD

human   AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
        *******************************************
chepo   AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
```

(57) Abstract

The present invention is directed to novel chimpanzee erythropoietin polypeptides and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL CHIMPANZEE ERYTHROPOIETIN (CHEPO) POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel chimpanzee erythropoietin polypeptides, nucleic acid molecules encoding those polypeptides and to the recombinant production of those polypeptides.

BACKGROUND OF THE INVENTION

Erythropoiesis, the production of red blood cells, occurs continuously throughout the human life span to offset cell destruction. Erythropoiesis is a very precisely controlled physiological mechanism enabling sufficient numbers of red blood cells to be available in the blood for proper tissue oxygenation, but not so many that the cells would impede circulation. The formation of red blood cells occurs in the bone marrow and is under control of the hormone, erythropoietin.

Erythropoietin, an acidic glycoprotein is approximately 34,000 dalton molecular weight, may occur in three forms: alpha, beta and asialo. The alpha and beta forms different slightly in carbohydrate components have the same potency, biological activity and molecular weight. The asialo form is an alpha or beta form with the terminal carbohydrate (sialic acid) removed. Erythropoietin is present in a very low concentrations in plasma when the body is in a healthy state wherein tissues receive sufficient oxygenation from the existing number of erythrocytes. Tbis normal low concentration is enough to stimulate replacement of red blood cells which are lost normally through aging.

The amount of erythropoietin in the circulation is increased under conditions of hypoxia when oxygen transport by blood cells in the circulation is reduced. Hypoxia may be caused by loss of large amounts of blood through hemorrhage, destruction of red blood cells by over-exposure to radiation, reduction in oxygen intake due to high altitudes or prolonged unconsciousness, or various forms of anemia. In response to tissues undergoing hypoxic stress, erythropoietin will increase red blood cell production by stimulating the conversion of primitive precursor cells in the bone marrow into proerythroblasts which subsequently mature, synthesize hemoglobin and are released into the circulation as red blood cells. When the number of red blood cells in circulation is greater than needed for normal tissue oxygen requirements, erythropoietin in circulation is decreased.

Because erythropoietin is essential in the process of red blood cell formation, the hormone has potential useful application in both the diagnosis and treatment of blood disorders characterized by low or defective red blood cell production. See, generally, Pennathur-Das, et al., Blood 63(5):1168-71 (1984) and Haddy, Am. Jour. Ped. Hematol. Oncol., 4:191-196 (1982) relating to erythropoietin in possible therapies for sickle cell disease, and Eschbach et al., J. Clin. Invest. 74(2):434-441 (1984), describing a therapeutic regimen for uremic sheep based on *in vivo* response to erythropoietin-rich plasma infusions and proposing a dosage of 10 U EOP/kg per day for 15-40 days as corrective of anemia of the type associated with chronic renal failure. See also, Krane, Henry Ford Hosp. Med. J., 31(3):177-181 (1983).

We describe herein the identification and characterization of a novel erythropoietin polypeptide derived from the chimpanzee, designated herein as CHEPO .

SUMMARY OF THE INVENTION

A cDNA clone has been identified that has homology to nucleic acid encoding human erythropoietin that encodes a novel chimpanzee erythropoietin polypeptide, designated in the present application as "CHEPO".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a CHEPO polypeptide.

In one aspect, the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 80% nucleic acid sequence identity, alternatively at least about 81% nucleic acid sequence identity, alternatively at least about 82% nucleic acid sequence identity, alternatively at least about 83% nucleic acid sequence identity, alternatively at least about 84% nucleic acid sequence identity, alternatively at least about 85% nucleic acid sequence identity, alternatively at least about 86% nucleic acid sequence identity, alternatively at least about 87% nucleic acid sequence identity, alternatively at least about 88% nucleic acid sequence identity, alternatively at least about 89% nucleic acid sequence identity, alternatively at least about 90% nucleic acid sequence identity, alternatively at least about 91% nucleic acid sequence identity, alternatively at least about 92% nucleic acid sequence identity, alternatively at least about 93% nucleic acid sequence identity, alternatively at least about 94% nucleic acid sequence identity, alternatively at least about 95% nucleic acid sequence identity, alternatively at least about 96% nucleic acid sequence identity, alternatively at least about 97% nucleic acid sequence identity, alternatively at least about 98% nucleic acid sequence identity and alternatively at least about 99% nucleic acid sequence identity to (a) a DNA molecule encoding a CHEPO polypeptide having the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or (b) the complement of the DNA molecule of (a).

In another aspect, the isolated nucleic acid molecule comprises (a) a nucleotide sequence encoding a CHEPO polypeptide having the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or (b) the complement of the nucleotide sequence of (a).

In other aspects, the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 80% nucleic acid sequence identity, alternatively at least about 81% nucleic acid sequence identity, alternatively at least about 82% nucleic acid sequence identity, alternatively at least about 83% nucleic acid sequence identity, alternatively at least about 84% nucleic acid sequence identity, alternatively at least about 85% nucleic acid sequence identity, alternatively at least about 86% nucleic acid sequence identity, alternatively at least about 87% nucleic acid sequence identity, alternatively at least about 88% nucleic acid sequence identity, alternatively at least about 89% nucleic acid sequence identity, alternatively at least about 90% nucleic acid sequence identity, alternatively at least about 91% nucleic acid sequence identity, alternatively at least about 92% nucleic acid sequence identity, alternatively at least about 93% nucleic acid sequence identity, alternatively at least about 94% nucleic acid sequence identity, alternatively at least about 95% nucleic acid sequence identity, alternatively at least about 96% nucleic acid sequence identity, alternatively at least about 97% nucleic acid sequence identity, alternatively at least about 98% nucleic acid sequence identity and alternatively at least about 99% nucleic acid sequence identity to (a) a DNA molecule having the sequence of nucleotides from about 1 or about 82 to about 579, inclusive, of Figure 2 (SEQ ID NO:3), or (b) the complement of the DNA molecule of (a).

In another aspect, the isolated nucleic acid molecule comprises (a) the nucleotide sequence of from about 1 or about 82 to about 579, inclusive, of Figure 2 (SEQ ID NO:3), or (b) the complement of the nucleotide sequence of (a).

In another aspect, the invention concerns an isolated nucleic acid molecule which encodes an active CHEPO polypeptide as defined below comprising a nucleotide sequence that hybridizes to the complement of a nucleic acid sequence that encodes amino acids 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5). Preferably, hybridization occurs under stringent hybridization and wash conditions.

In yet another aspect, the invention concerns an isolated nucleic acid molecule which encodes an active CHEPO polypeptide as defined below comprising a nucleotide sequence that hybridizes to the complement of the nucleic acid sequence between about nucleotides 1 or about 82 and about 579, inclusive, of Figure 2 (SEQ ID NO:3). Preferably, hybridization occurs under stringent hybridization and wash conditions.

In a further aspect, the invention concerns an isolated nucleic acid molecule which is produced by hybridizing a test DNA molecule under stringent conditions with (a) a DNA molecule encoding a CHEPO polypeptide having the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or (b) the complement of the DNA molecule of (a), and, if the test DNA molecule has at least about an 80% nucleic acid sequence identity, alternatively at least about 81% nucleic acid sequence identity, alternatively at least about 82% nucleic acid sequence identity, alternatively at least about 83% nucleic acid sequence identity, alternatively at least about 84% nucleic acid sequence identity, alternatively at least about 85% nucleic acid sequence identity, alternatively at least about 86% nucleic acid sequence identity, alternatively at least about 87% nucleic acid sequence identity, alternatively at least about 88% nucleic acid sequence identity, alternatively at least about 89% nucleic acid sequence identity, alternatively at least about 90% nucleic acid sequence identity, alternatively at least about 91% nucleic acid sequence identity, alternatively at least about 92% nucleic acid sequence identity, alternatively at least about 93% nucleic acid sequence identity, alternatively at least about 94% nucleic acid sequence identity, alternatively at least about 95% nucleic acid sequence identity, alternatively at least about 96% nucleic acid sequence identity, alternatively at least about 97% nucleic acid sequence identity, alternatively at least about 98% nucleic acid sequence identity and alternatively at least about 99% nucleic acid sequence identity to (a) or (b), and isolating the test DNA molecule.

In another aspect, the invention concerns an isolated nucleic acid molecule comprising (a) a nucleotide sequence encoding a polypeptide scoring at least about 80% positives, alternatively at least about 81% positives, alternatively at least about 82% positives, alternatively at least about 83% positives, alternatively at least about 84% positives, alternatively at least about 85% positives, alternatively at least about 86% positives, alternatively at least about 87% positives, alternatively at least about 88% positives, alternatively at least about 89% positives, alternatively at least about 90% positives, alternatively at least about 91% positives, alternatively at least about 92% positives, alternatively at least about 93% positives, alternatively at least about 94% positives, alternatively at least about 95% positives, alternatively at least about 96% positives, alternatively at least about 97% positives, alternatively at least about 98% positives and alternatively at least about 99% positives when compared with the amino acid sequence of residues about 1 or about 28 to 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or (b) the complement of the nucleotide sequence of (a).

In a specific aspect, the invention provides an isolated nucleic acid molecule comprising DNA encoding a CHEPO polypeptide without the N-terminal signal sequence and/or the initiating methionine, or is complementary to such encoding nucleic acid molecule. The signal peptide has been tentatively identified as extending from about amino acid position 1 to about amino acid position 27 in the sequence of Figure 3 (SEQ ID NOS:2 and 5). It is noted, however, that the C-terminal boundary of the signal peptide may vary, but most likely by no more than about 5 amino acids on either side of the signal peptide C-terminal boundary as initially identified herein, wherein the C-terminal boundary of the signal peptide may be identified pursuant to criteria routinely employed in the art for identifying that type of amino acid sequence element (e.g., Nielsen et al., Prot. Eng. 10:1-6 (1997) and von Heinje et al., Nucl. Acids. Res. 14:4683-4690 (1986)). Moreover, it is also recognized that, in some cases, cleavage of a signal sequence from a secreted polypeptide is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding them, are contemplated by the present invention. As such, for purposes of the present application, the signal peptide of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5) extends from amino acids 1 to X of Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5). Therefore, mature forms of the CHEPO polypeptide which are encompassed by the present invention include those comprising amino acids X to 193 of Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5) and variants thereof as described below. Isolated nucleic acid molecules encoding these polypeptides are also contemplated.

Another embodiment is directed to fragments of a CHEPO polypeptide coding sequence that may find use as, for example, hybridization probes or for encoding fragments of a CHEPO polypeptide that may optionally encode a polypeptide comprising a binding site for an anti-CHEPO antibody. Such nucleic acid fragments are usually at least about 20 nucleotides in length, alternatively at least about 30 nucleotides in length, alternatively at least about 40 nucleotides in length, alternatively at least about 50 nucleotides in length, alternatively at least about 60 nucleotides in length, alternatively at least about 70 nucleotides in length, alternatively at least about 80 nucleotides in length, alternatively at least about 90 nucleotides in length, alternatively at least about 100 nucleotides in length, alternatively at least about 110 nucleotides in length, alternatively at least about 120 nucleotides in length, alternatively at least about 130 nucleotides in length, alternatively at least about 140 nucleotides in length, alternatively at least about 150 nucleotides in length, alternatively at least about 160 nucleotides in length, alternatively at least about 170 nucleotides in length, alternatively at least about 180 nucleotides in length, alternatively at least about 190 nucleotides in length, alternatively at least about 200 nucleotides in length, alternatively at least about 250 nucleotides in length, alternatively at least about 300 nucleotides in length, alternatively at least about 350 nucleotides in length, alternatively at least about 400 nucleotides in length, alternatively at least about 450 nucleotides in length, alternatively at least about 500 nucleotides in length, alternatively at least about 600 nucleotides in length, alternatively at least about 700 nucleotides in length, alternatively at least about 800 nucleotides in length, alternatively at least about 900 nucleotides in length and alternatively at least about 1000 nucleotides in length, wherein in this context the term about means the referenced nucleotide sequence length plus or minus 10% of that referenced length. In a preferred embodiment, the nucleotide sequence fragment is derived from any coding region of the nucleotide sequence shown in Figure 1 (SEQ ID NO:1). It is noted that novel fragments of a CHEPO polypeptide-encoding nucleotide sequence may be determined in a routine manner by aligning the CHEPO

polypeptide-encoding nucleotide sequence with other known nucleotide sequences using any of a number of well known sequence alignment programs and determining which CHEPO polypeptide-encoding nucleotide sequence fragment(s) are novel. All of such CHEPO polypeptide-encoding nucleotide sequences are contemplated herein and can be determined without undue experimentation. Also contemplated are the CHEPO polypeptide fragments encoded by these nucleotide molecule fragments, preferably those CHEPO polypeptide fragments that comprise a binding site for an anti-CHEPO antibody.

In another embodiment, the invention provides a vector comprising a nucleotide sequence encoding CHEPO or its variants. The vector may comprise any of the isolated nucleic acid molecules hereinabove identified.

A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing CHEPO polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of CHEPO and recovering CHEPO from the cell culture.

In another embodiment, the invention provides isolated CHEPO polypeptide encoded by any of the isolated nucleic acid sequences hereinabove identified.

In a specific aspect, the invention provides isolated native sequence CHEPO polypeptide, which in certain embodiments, includes an amino acid sequence comprising residues from about 1 or about 28 to about 193 of Figure 3 (SEQ ID NOS:2 and 5).

In another aspect, the invention concerns an isolated CHEPO polypeptide, comprising an amino acid sequence having at least about 80% amino acid sequence identity, alternatively at least about 81% amino acid sequence identity, alternatively at least about 82% amino acid sequence identity, alternatively at least about 83% amino acid sequence identity, alternatively at least about 84% amino acid sequence identity, alternatively at least about 85% amino acid sequence identity, alternatively at least about 86% amino acid sequence identity, alternatively at least about 87% amino acid sequence identity, alternatively at least about 88% amino acid sequence identity, alternatively at least about 89% amino acid sequence identity, alternatively at least about 90% amino acid sequence identity, alternatively at least about 91% amino acid sequence identity, alternatively at least about 92% amino acid sequence identity, alternatively at least about 93% amino acid sequence identity, alternatively at least about 94% amino acid sequence identity, alternatively at least about 95% amino acid sequence identity, alternatively at least about 96% amino acid sequence identity, alternatively at least about 97% amino acid sequence identity, alternatively at least about 98% amino acid sequence identity and alternatively at least about 99% amino acid sequence identity to the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5).

In a further aspect, the invention concerns an isolated CHEPO polypeptide comprising an amino acid sequence scoring at least about 80% positives, alternatively at least about 81% positives, alternatively at least about 82% positives, alternatively at least about 83% positives, alternatively at least about 84% positives, alternatively at least about 85% positives, alternatively at least about 86% positives, alternatively at least about 87% positives, alternatively at least about 88% positives, alternatively at least about 89% positives, alternatively at least about 90% positives, alternatively at least about 91% positives, alternatively at least about 92% positives, alternatively at least about 93% positives, alternatively at least about 94% positives, alternatively at least about 95% positives, alternatively at least about 96% positives,

alternatively at least about 97% positives, alternatively at least about 98% positives and alternatively at least about 99% positives when compared with the amino acid sequence of residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5).

In a specific aspect, the invention provides an isolated CHEPO polypeptide without the N-terminal signal sequence and/or the initiating methionine and is encoded by a nucleotide sequence that encodes such an amino acid sequence as hereinbefore described. Processes for producing the same are also herein described, wherein those processes comprise culturing a host cell comprising a vector which comprises the appropriate encoding nucleic acid molecule under conditions suitable for expression of the CHEPO polypeptide and recovering the CHEPO polypeptide from the cell culture.

In yet another aspect, the invention concerns an isolated CHEPO polypeptide, comprising the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or a fragment thereof which is biologically active or sufficient to provide a binding site for an anti-CHEPO antibody, wherein the identification of CHEPO polypeptide fragments that possess biological activity or provide a binding site for an anti-CHEPO antibody may be accomplished in a routine manner using techniques which are well known in the art. Preferably, the CHEPO fragment retains a qualitative biological activity of a native CHEPO polypeptide.

In a still further aspect, the invention provides a polypeptide produced by (i) hybridizing a test DNA molecule under stringent conditions with (a) a DNA molecule encoding a CHEPO polypeptide having the sequence of amino acid residues from about 1 or about 28 to about 193, inclusive, of Figure 3 (SEQ ID NOS:2 and 5), or (b) the complement of the DNA molecule of (a), and if the test DNA molecule has at least about an 80% nucleic acid sequence identity, alternatively at least about 81% nucleic acid sequence identity, alternatively at least about 82% nucleic acid sequence identity, alternatively at least about 83% nucleic acid sequence identity, alternatively at least about 84% nucleic acid sequence identity, alternatively at least about 85% nucleic acid sequence identity, alternatively at least about 86% nucleic acid sequence identity, alternatively at least about 87% nucleic acid sequence identity, alternatively at least about 88% nucleic acid sequence identity, alternatively at least about 89% nucleic acid sequence identity, alternatively at least about 90% nucleic acid sequence identity, alternatively at least about 91% nucleic acid sequence identity, alternatively at least about 92% nucleic acid sequence identity, alternatively at least about 93% nucleic acid sequence identity, alternatively at least about 94% nucleic acid sequence identity, alternatively at least about 95% nucleic acid sequence identity, alternatively at least about 96% nucleic acid sequence identity, alternatively at least about 97% nucleic acid sequence identity, alternatively at least about 98% nucleic acid sequence identity and alternatively at least about 99% nucleic acid sequence identity to (a) or (b), (ii) culturing a host cell comprising the test DNA molecule under conditions suitable for expression of the polypeptide, and (iii) recovering the polypeptide from the cell culture.

In another embodiment, the invention provides chimeric molecules comprising a CHEPO polypeptide fused to a heterologous polypeptide or amino acid sequence, wherein the CHEPO polypeptide may comprise any CHEPO polypeptide, variant or fragment thereof as hereinbefore described. An example of such a chimeric molecule comprises a CHEPO polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody as defined below which specifically binds to a CHEPO polypeptide as hereinbefore described. Optionally, the antibody is a monoclonal antibody, an antibody fragment or a single chain antibody.

In yet another embodiment, the invention concerns agonists and antagonists of a native CHEPO polypeptide as defined below. In a particular embodiment, the agonist or antagonist is an anti-CHEPO antibody or a small molecule.

In a further embodiment, the invention concerns a method of identifying agonists or antagonists to a CHEPO polypeptide which comprise contacting the CHEPO polypeptide with a candidate molecule and monitoring a biological activity mediated by said CHEPO polypeptide. Preferably, the CHEPO polypeptide is a native CHEPO polypeptide.

In a still further embodiment, the invention concerns a composition of matter comprising a CHEPO polypeptide, or an agonist or antagonist of a CHEPO polypeptide as herein described, or an anti-CHEPO antibody, in combination with a carrier. Optionally, the carrier is a pharmaceutically acceptable carrier.

Another embodiment of the present invention is directed to the use of a CHEPO polypeptide, or an agonist or antagonist thereof as herein described, or an anti-CHEPO antibody, for the preparation of a medicament useful in the treatment of a condition which is responsive to the CHEPO polypeptide, an agonist or antagonist thereof or an anti-CHEPO antibody.

Yet another embodiment of the present invention is directed to CHEPO polypeptides having altered glycosylation patterns in one or more regions of the polypeptide as compared to the native sequence CHEPO polypeptide, preferably in the region surrounding and/or including amino acid position 84 in the CHEPO amino acids sequence shown in Figure 3 (SEQ ID NOS:2 and 5). In various embodiments, CHEPO variant polypeptides are prepared using well known techniques so as to create an N- or O-linked glycosylation site at or near amino acid position 84 in the CHEPO polypeptide sequence. For example, CHEPO polypeptides contemplated by the present invention include those where (a) amino acids 81-84 of the CHEPO amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5) (i.e., Met-Glu-Val-Arg; SEQ ID NO:6) are replaced by the amino acid sequence Asn-X-Ser-X (SEQ ID NO:7) or Asn-X-Thr-X (SEQ ID NO:8), where X is any amino acid except for Pro; (b) amino acids 82-85 of the CHEPO amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5) (i.e., Glu-Val-Arg-Gln; SEQ ID NO:9) are replaced by the amino acid sequence Asn-X-Ser-X (SEQ ID NO:7) or Asn-X-Thr-X (SEQ ID NO:8), where X is any amino acid except for Pro; (c) amino acids 83-86 of the CHEPO amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5) (i.e., Val-Arg-Gln-Gln; SEQ ID NO:10) are replaced by the amino acid sequence Asn-X-Ser-X (SEQ ID NO:7) or Asn-X-Thr-X (SEQ ID NO:8), where X is any amino acid except for Pro; or (d) amino acids 84-87 of the CHEPO amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5) (i.e., Arg-Gln-Gln-Ala; SEQ ID NO:11) are replaced by the amino acid sequence Asn-X-Ser-X (SEQ ID NO:7) or Asn-X-Thr-X (SEQ ID NO:8), where X is any amino acid except for Pro, thereby creating an N-glycosylation site at those positions. Nucleic acids encoding these variant polypeptides are also contemplated herein as are vectors and host cells comprising those nucleic acids.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-C show a nucleotide sequence (SEQ ID NO:1) of an isolated genomic DNA molecule containing a nucleotide sequence (nucleotides 134-146, 667-812, 1071-1157, 1760-1939 and 2074-2226, exclusive of others)

encoding native sequence CHEPO. Also presented in the genomic sequence are the locations of the start codon, exons and introns as well as the amino acid sequence (SEQ ID NO:2) encoded by the coding sequence of SEQ ID NO:1.

Figure 2 shows the cDNA sequence of the CHEPO molecule (SEQ ID NO:3) and the amino acid sequence encoded thereby (SEQ ID NO:2).

Figure 3 shows a comparison of the human erythropoietin amino acid sequence (human) (SEQ ID NO:4) and that of the chimp erythropoietin (chepo) described herein, wherein the amino acid designated X at amino acid position 142 of the CHEPO sequence is either glutamine (SEQ ID NO:2) or lysine (SEQ ID NO:5).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "CHEPO polypeptide", "CHEPO protein" and "CHEPO" when used herein encompass native sequence CHEPO and CHEPO polypeptide variants (which are further defined herein). The CHEPO polypeptide may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant and/or synthetic methods.

A "native sequence CHEPO" comprises a polypeptide having the same amino acid sequence as a CHEPO derived from nature. Such native sequence CHEPO can be isolated from nature or can be produced by recombinant and/or synthetic means. The term "native sequence CHEPO" specifically encompasses naturally-occurring truncated or secreted forms (*e.g.*, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of the CHEPO. In one embodiment of the invention, the native sequence CHEPO is a mature or full-length native sequence CHEPO comprising amino acids 1 to 193 of Figure 3 (SEQ ID NOS:2 and 5). Also, while the CHEPO polypeptides disclosed in Figure 3 (SEQ ID NOS:2 and 5) is shown to begin with the methionine residue designated herein as amino acid position 1, it is conceivable and possible that another methionine residue located either upstream or downstream from amino acid position 1 in Figure 3 (SEQ ID NOS:2 and 5) may be employed as the starting amino acid residue for the CHEPO polypeptide.

"CHEPO variant polypeptide" means an active CHEPO polypeptide as defined below having at least about 80% amino acid sequence identity with the amino acid sequence of (a) residues 1 or about 28 to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), (b) X to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid residue from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5) or (c) another specifically derived fragment of the amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5). Such CHEPO variant polypeptides include, for instance, CHEPO polypeptides wherein one or more amino acid residues are added, or deleted, at the N- and/or C-terminus, as well as within one or more internal domains, of the sequence of Figure 3 (SEQ ID NOS:2 and 5). Ordinarily, a CHEPO variant polypeptide will have at least about 80% amino acid sequence identity, alternatively at least about 81% amino acid sequence identity, alternatively at least about 82% amino acid sequence identity, alternatively at least about 83% amino acid sequence identity, alternatively at least about 84% amino acid sequence identity, alternatively at least about 85% amino acid sequence identity, alternatively at least about 86% amino acid sequence identity, alternatively at least about 87% amino acid sequence identity, alternatively at least about 88% amino acid sequence identity, alternatively

at least about 89% amino acid sequence identity, alternatively at least about 90% amino acid sequence identity, alternatively at least about 91% amino acid sequence identity, alternatively at least about 92% amino acid sequence identity, alternatively at least about 93% amino acid sequence identity, alternatively at least about 94% amino acid sequence identity, alternatively at least about 95% amino acid sequence identity, alternatively at least about 96% amino acid sequence identity, alternatively at least about 97% amino acid sequence identity, alternatively at least about 98% amino acid sequence identity and alternatively at least about 99% amino acid sequence identity with (a) residues 1 or about 28 to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), (b) X to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid residue from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5) or (c) another specifically derived fragment of the amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5). CHEPO variant polypeptides do not encompass the native CHEPO polypeptide sequence. Ordinarily, CHEPO variant polypeptides are at least about 10 amino acids in length, alternatively at least about 20 amino acids in length, alternatively at least about 30 amino acids in length, alternatively at least about 40 amino acids in length, alternatively at least about 50 amino acids in length, alternatively at least about 60 amino acids in length, alternatively at least about 70 amino acids in length, alternatively at least about 80 amino acids in length, alternatively at least about 90 amino acids in length, alternatively at least about 100 amino acids in length, alternatively at least about 150 amino acids in length, alternatively at least about 200 amino acids in length, alternatively at least about 300 amino acids in length, or more.

"Percent (%) amino acid sequence identity" with respect to the CHEPO polypeptide sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a CHEPO sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % amino acid sequence identity values are obtained as described below by using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 1 below. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Table 1 has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 1. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

For purposes herein, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

100 times the fraction X/Y

where X is the number of amino acid residues scored as identical matches by the sequence alignment program ALIGN-2 in that program s alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A. As examples of % amino acid sequence identity calculations, Tables 2 and 3 below demonstrate how to calculate the % amino acid sequence identity of the amino acid sequence designated Comparison Protein to the amino acid sequence designated PRO .

Unless specifically stated otherwise, all % amino acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer program. However, % amino acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from http://www.ncbi.nlm.nih.gov or otherwise obtained from the National Institute of Health, Bethesda, MD. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

100 times the fraction X/Y

where X is the number of amino acid residues scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program s alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A.

"CHEPO variant polynucleotide" or CHEPO variant nucleic acid sequence means a nucleic acid molecule which encodes an active CHEPO polypeptide as defined below and which has at least about 80% nucleic acid sequence identity with either (a) a nucleic acid sequence which encodes residues 1 or about 28 to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), (b) a nucleic acid sequence which encodes residues X to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid residue from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5) or (c) a nucleic acid sequence which encodes another specifically derived fragment of the amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5). Ordinarily, a CHEPO variant polynucleotide will have at least about 80% nucleic acid sequence identity, alternatively at least about 81% nucleic acid sequence identity, alternatively at least about 82% nucleic acid sequence identity, alternatively at least about 83% nucleic acid sequence identity, alternatively at least about

84% nucleic acid sequence identity, alternatively at least about 85% nucleic acid sequence identity, alternatively at least about 86% nucleic acid sequence identity, alternatively at least about 87% nucleic acid sequence identity, alternatively at least about 88% nucleic acid sequence identity, alternatively at least about 89% nucleic acid sequence identity, alternatively at least about 90% nucleic acid sequence identity, alternatively at least about 91% nucleic acid sequence identity, alternatively at least about 92% nucleic acid sequence identity, alternatively at least about 93% nucleic acid sequence identity, alternatively at least about 94% nucleic acid sequence identity, alternatively at least about 95% nucleic acid sequence identity, alternatively at least about 96% nucleic acid sequence identity, alternatively at least about 97% nucleic acid sequence identity, alternatively at least about 98% nucleic acid sequence identity and alternatively at least about 99% nucleic acid sequence identity with either (a) a nucleic acid sequence which encodes residues 1 or about 28 to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), (b) a nucleic acid sequence which encodes residues X to 193 of the CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5), wherein X is any amino acid residue from 23 to 32 of Figure 3 (SEQ ID NOS:2 and 5) or (c) a nucleic acid sequence which encodes another specifically derived fragment of the amino acid sequence shown in Figure 3 (SEQ ID NOS:2 and 5). CHEPO polynucleotide variants do not encompass the native CHEPO nucleotide sequence.

Ordinarily, CHEPO variant polynucleotides are at least about 30 nucleotides in length, alternatively at least about 60 nucleotides in length, alternatively at least about 90 nucleotides in length, alternatively at least about 120 nucleotides in length, alternatively at least about 150 nucleotides in length, alternatively at least about 180 nucleotides in length, alternatively at least about 210 nucleotides in length, alternatively at least about 240 nucleotides in length, alternatively at least about 270 nucleotides in length, alternatively at least about 300 nucleotides in length, alternatively at least about 450 nucleotides in length, alternatively at least about 600 nucleotides in length, alternatively at least about 900 nucleotides in length, or more.

"Percent (%) nucleic acid sequence identity" with respect to the CHEPO polypeptide-encoding nucleic acid sequences identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in a CHEPO polypeptide-encoding nucleic acid sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % nucleic acid sequence identity values are obtained as described below by using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 1. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Table 1 has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 1. The ALIGN-2 program should be compiled for use on a UNIX

operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

For purposes herein, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

100 times the fraction W/Z

where W is the number of nucleotides scored as identical matches by the sequence alignment program ALIGN-2 in that program s alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C. As examples of % nucleic acid sequence identity calculations, Tables 4 and 5 below demonstrate how to calculate the % nucleic acid sequence identity of the nucleic acid sequence designated Comparison DNA to the nucleic acid sequence designated PRO-DNA .

Unless specifically stated otherwise, all % nucleic acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer program. However, % nucleic acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from http://www.ncbi.nlm.nih.gov or otherwise obtained from the National Institute of Health, Bethesda, MD. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

100 times the fraction W/Z

where W is the number of nucleotides scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program s alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C.

In other embodiments, CHEPO variant polynucleotides are nucleic acid molecules that encode an active CHEPO polypeptide and which are capable of hybridizing, preferably under stringent hybridization and wash conditions, to

nucleotide sequences encoding the full-length CHEPO polypeptide shown in Figure 3 (SEQ ID NOS:2 and 5). CHEPO variant polypeptides may be those that are encoded by a CHEPO variant polynucleotide.

The term "positives", in the context of the amino acid sequence identity comparisons performed as described above, includes amino acid residues in the sequences compared that are not only identical, but also those that have similar properties. Amino acid residues that score a positive value to an amino acid residue of interest are those that are either identical to the amino acid residue of interest or are a preferred substitution (as defined in Table 6 below) of the amino acid residue of interest.

For purposes herein, the % value of positives of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % positives to, with, or against a given amino acid sequence B) is calculated as follows:

100 times the fraction X/Y

where X is the number of amino acid residues scoring a positive value as defined above by the sequence alignment program ALIGN-2 in that program s alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % positives of A to B will not equal the % positives of B to A.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Preferably, the isolated polypeptide is free of association with all components with which it is naturally associated. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the CHEPO natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" nucleic acid molecule encoding a CHEPO polypeptide is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the CHEPO-encoding nucleic acid. Preferably, the isolated nucleic is free of association with all components with which it is naturally associated. An isolated CHEPO-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated nucleic acid molecules therefore are distinguished from the CHEPO-encoding nucleic acid molecule as it exists in natural cells. However, an isolated nucleic acid molecule encoding a CHEPO polypeptide includes CHEPO-encoding nucleic acid molecules contained in cells that ordinarily express CHEPO where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers, for example, single anti-CHEPO monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies), anti-CHEPO antibody compositions with polyepitopic specificity, single chain anti-CHEPO antibodies, and fragments of anti-CHEPO antibodies (see below). The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50 C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42 C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt s solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42 C, with washes at 42 C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55 C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55 C.

"Moderately stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and %SDS) less stringent that those described above. An example of moderately stringent conditions is overnight incubation at 37 C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt s solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50 C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising a CHEPO polypeptide fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere with activity of the polypeptide to which it is fused. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 and 50 amino acid residues (preferably, between about 10 and 20 amino acid residues).

As used herein, the term "immunoadhesin" designates antibody-like molecules which combine the binding specificity of a heterologous protein (an "adhesin") with the effector functions of immunoglobulin constant domains. Structurally, the immunoadhesins comprise a fusion of an amino acid sequence with the desired binding specificity which is other than the antigen recognition and binding site of an antibody (i.e., is "heterologous"), and an immunoglobulin constant domain sequence. The adhesin part of an immunoadhesin molecule typically is a contiguous amino acid sequence comprising at least the binding site of a receptor or a ligand. The immunoglobulin constant domain sequence in the immunoadhesin may be obtained from any immunoglobulin, such as IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA (including IgA-1 and IgA-2), IgE, IgD or IgM.

"Active" or "activity" for the purposes herein refers to form(s) of CHEPO which retain a biological and/or an immunological activity of native or naturally-occurring CHEPO, wherein biological activity refers to a biological function (either inhibitory or stimulatory) caused by a native or naturally-occurring CHEPO other than the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring CHEPO and an immunological activity refers to the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring CHEPO. Preferred biological activities includes, for example, the ability to regulate red blood cell production, to bind to receptors on the surface of committed progenitor cells of the bone marrow and/or other hematopoietic tissues and/or to induce proliferation and/or terminal maturation of erythroid cells.

The term "antagonist" is used in the broadest sense, and includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native CHEPO polypeptide disclosed herein. In a similar manner, the term "agonist" is used in the broadest sense and includes any molecule that mimics a biological activity of a native CHEPO polypeptide disclosed herein. Suitable agonist or antagonist molecules specifically include agonist or antagonist antibodies or antibody fragments, fragments or amino acid sequence variants of native CHEPO polypeptides, peptides, small organic molecules, etc. Methods for identifying agonists or antagonists of a CHEPO polypeptide may comprise contacting a

CHEPO polypeptide with a candidate agonist or antagonist molecule and measuring a detectable change in one or more biological activities normally associated with the CHEPO polypeptide.

"Treatment" refers to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) the targeted pathologic condition or disorder. Those in need of treatment include those already with the disorder as well as those prone to have the disorder or those in whom the disorder is to be prevented.

"Chronic" administration refers to administration of the agent(s) in a continuous mode as opposed to an acute mode, so as to maintain the initial therapeutic effect (activity) for an extended period of time. Intermittent administration is treatment that is not consecutively done without interruption, but rather is cyclic in nature.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, cats, cattle, horses, sheep, pigs, goats, rabbits, etc. Preferably, the mammal is human.

Administration "in combination with" one or more further therapeutic agents includes simultaneous (concurrent) and consecutive administration in any order.

"Carriers" as used herein include pharmaceutically acceptable carriers, excipients, or stabilizers which are nontoxic to the cell or mammal being exposed thereto at the dosages and concentrations employed. Often the physiologically acceptable carrier is an aqueous pH buffered solution. Examples of physiologically acceptable carriers include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptide; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN , polyethylene glycol (PEG), and PLURONICS .

"Antibody fragments" comprise a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', $F(ab')_2$, and Fv fragments; diabodies; linear antibodies (Zapata et al., Protein Eng. 8(10): 1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, a designation reflecting the ability to crystallize readily. Pepsin treatment yields an $F(ab')_2$ fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

"Fv" is the minimum antibody fragment which contains a complete antigen-recognition and -binding site. This region consists of a dimer of one heavy- and one light-chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V_H-V_L dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. $F(ab')_2$ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa and lambda, based on the amino acid sequences of their constant domains.

Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA, and IgA2.

"Single-chain Fv" or "sFv" antibody fragments comprise the VH and VL domains of antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the VH and VL domains which enables the sFv to form the desired structure for antigen binding. For a review of sFv, see Pluckthun in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (VH) connected to a light-chain variable domain (VL) in the same polypeptide chain (VH - VL). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger et al., Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993).

An "isolated" antibody is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified (1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody in situ within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

The word "label" when used herein refers to a detectable compound or composition which is conjugated directly or indirectly to the antibody so as to generate a "labeled" antibody. The label may be detectable by itself (e.g. radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition which is detectable.

By "solid phase" is meant a non-aqueous matrix to which the antibody of the present invention can adhere. Examples of solid phases encompassed herein include those formed partially or entirely of glass (e.g., controlled pore glass), polysaccharides (e.g., agarose), polyacrylamides, polystyrene, polyvinyl alcohol and silicones. In certain embodiments, depending on the context, the solid phase can comprise the well of an assay plate; in others it is a purification column (e.g., an affinity chromatography column). This term also includes a discontinuous solid phase of discrete particles, such as those described in U.S. Patent No. 4,275,149.

A "liposome" is a small vesicle composed of various types of lipids, phospholipids and/or surfactant which is useful for delivery of a drug (such as a CHEPO polypeptide or antibody thereto) to a mammal. The components of the liposome are commonly arranged in a bilayer formation, similar to the lipid arrangement of biological membranes.

A small molecule is defined herein to have a molecular weight below about 500 Daltons.

Table 1

```
/*
 *
 * C-C increased from 12 to 15
 * Z is average of EQ
 * B is average of ND
 * match with stop is _M; stop-stop = 0; J (joker) match = 0
 */
#define _M      -8      /* value of a match with a stop */

int     _day[26][26] = {
/*    A B C D E F G H I J K L M N O P Q R S T U V W X Y Z */
/* A */  { 2, 0,-2, 0, 0,-4, 1,-1,-1, 0,-1,-2,-1, 0,_M, 1, 0,-2, 1, 1, 0, 0,-6, 0,-3, 0},
/* B */  { 0, 3,-4, 3, 2,-5, 0, 1,-2, 0, 0,-3,-2, 2,_M,-1, 1, 0, 0, 0, 0,-2,-5, 0,-3, 1},
/* C */  {-2,-4,15,-5,-5,-4,-3,-3,-2, 0,-5,-6,-5,-4,_M,-3,-5,-4, 0,-2, 0,-2,-8, 0, 0,-5},
/* D */  { 0, 3,-5, 4, 3,-6, 1, 1,-2, 0, 0,-4,-3, 2,_M,-1, 2,-1, 0, 0, 0,-2,-7, 0,-4, 2},
/* E */  { 0, 2,-5, 3, 4,-5, 0, 1,-2, 0, 0,-3,-2, 1,_M,-1, 2,-1, 0, 0, 0,-2,-7, 0,-4, 3},
/* F */  {-4,-5,-4,-6,-5, 9,-5,-2, 1, 0,-5, 2, 0,-4,_M,-5,-5,-4,-3,-3, 0,-1, 0, 0, 7,-5},
/* G */  { 1, 0,-3, 1, 0,-5, 5,-2,-3, 0,-2,-4,-3, 0,_M,-1,-1,-3, 1, 0, 0,-1,-7, 0,-5, 0},
/* H */  {-1, 1,-3, 1, 1,-2,-2, 6,-2, 0, 0,-2,-2, 2,_M, 0, 3, 2,-1,-1, 0,-2,-3, 0, 0, 2},
/* I */  {-1,-2,-2,-2,-2, 1,-3,-2, 5, 0,-2, 2, 2,-2,_M,-2,-2,-2,-1, 0, 0, 4,-5, 0,-1,-2},
/* J */  { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,_M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
/* K */  {-1, 0,-5, 0, 0,-5,-2, 0,-2, 0, 5,-3, 0, 1,_M,-1, 1, 3, 0, 0, 0,-2,-3, 0,-4, 0},
/* L */  {-2,-3,-6,-4,-3, 2,-4,-2, 2, 0,-3, 6, 4,-3,_M,-3,-2,-3,-3,-1, 0, 2,-2, 0,-1,-2},
/* M */  {-1,-2,-5,-3,-2, 0,-3,-2, 2, 0, 0, 4, 6,-2,_M,-2,-1, 0,-2,-1, 0, 2,-4, 0,-2,-1},
/* N */  { 0, 2,-4, 2, 1,-4, 0, 2,-2, 0, 1,-3,-2, 2,_M,-1, 1, 0, 1, 0, 0,-2,-4, 0,-2, 1},
/* O */  {_M,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M, 0,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M,_M},
/* P */  { 1,-1,-3,-1,-1,-5,-1, 0,-2, 0,-1,-3,-2,-1,_M, 6, 0, 0, 1, 0, 0,-1,-6, 0,-5, 0},
/* Q */  { 0, 1,-5, 2, 2,-5,-1, 3,-2, 0, 1,-2,-1, 1,_M, 0, 4, 1,-1,-1, 0,-2,-5, 0,-4, 3},
/* R */  {-2, 0,-4,-1,-1,-4,-3, 2,-2, 0, 3,-3, 0, 0,_M, 0, 1, 6, 0,-1, 0,-2, 2, 0,-4, 0},
/* S */  { 1, 0, 0, 0, 0,-3, 1,-1,-1, 0, 0,-3,-2, 1,_M, 1,-1, 0, 2, 1, 0,-1,-2, 0,-3, 0},
/* T */  { 1, 0,-2, 0, 0,-3, 0,-1, 0, 0, 0,-1,-1, 0,_M, 0,-1,-1, 1, 3, 0, 0,-5, 0,-3, 0},
/* U */  { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,_M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
/* V */  { 0,-2,-2,-2,-2,-1,-1,-2, 4, 0,-2, 2, 2,-2,_M,-1,-2,-2,-1, 0, 0, 4,-6, 0,-2,-2},
/* W */  {-6,-5,-8,-7,-7, 0,-7,-3,-5, 0,-3,-2,-4,-4,_M,-6,-5, 2,-2,-5, 0,-6,17, 0, 0,-6},
/* X */  { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,_M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
```

```
/* Y */    {-3,-3, 0,-4,-4, 7,-5, 0,-1, 0,-4,-1,-2,-2,_M,-5,-4,-4,-3,-3, 0,-2, 0, 0,10,-4},
/* Z */    { 0, 1,-5, 2, 3,-5, 0, 2,-2, 0, 0,-2,-1, 1,_M, 0, 3, 0, 0, 0, 0,-2,-6, 0,-4, 4}
};
```

Table 1 (cont)

```
/*
*/
#include <stdio.h>
#include <ctype.h>

#define MAXJMP      16      /* max jumps in a diag */
#define MAXGAP 24   /* don't continue to penalize gaps larger than this */
#define JMPS        1024    /* max jmps in an path */
#define MX          4       /* save if there's at least MX-1 bases since last jmp */

#define DMAT        3       /* value of matching bases */
#define DMIS        0       /* penalty for mismatched bases */
#define DINS0       8       /* penalty for a gap */
#define DINS1       1       /* penalty per base */
#define PINS0       8       /* penalty for a gap */
#define PINS1       4       /* penalty per residue */

struct jmp {
        short           n[MAXJMP];      /* size of jmp (neg for dely) */
        unsigned short  x[MAXJMP];      /* base no. of jmp in seq x */
};                                      /* limits seq to 2^16 -1 */

struct diag {
        int             score;          /* score at last jmp */
        long            offset;         /* offset of prev block */
        short           ijmp;           /* current jmp index */
        struct jmp      jp;             /* list of jmps */
};

struct path {
        int     spc;            /* number of leading spaces */
        short   n[JMPS]; /* size of jmp (gap) */
        int     x[JMPS]; /* loc of jmp (last elem before gap) */
};
```

```
char        *ofile;          /* output file name */
char        *namex[2];       /* seq names: getseqs() */
char        *prog;           /* prog name for err msgs */
char        *seqx[2];        /* seqs: getseqs() */
int         dmax;            /* best diag: nw() */
int         dmax0;           /* final diag */
int         dna;             /* set if dna: main() */
int         endgaps;         /* set if penalizing end gaps */
int         gapx, gapy;      /* total gaps in seqs */
int         len0, len1;      /* seq lens */
int         ngapx, ngapy;    /* total size of gaps */
int         smax;            /* max score: nw() */
int         *xbm;            /* bitmap for matching */
long        offset;          /* current offset in jmp file */
struct diag *dx;             /* holds diagonals */
struct path pp[2];           /* holds path for seqs */

char        *calloc(), *malloc(), *index(), *strcpy();
char        *getseq(), *g_calloc();
```

Table 1 (cont.)

```
/* Needleman-Wunsch alignment program
 *
 * usage: progs file1 file2
 *   where file1 and file2 are two dna or two protein sequences.
 *   The sequences can be in upper- or lower-case an may contain ambiguity
 *   Any lines beginning with ';', '>' or '<' are ignored
 *   Max file length is 65535 (limited by unsigned short x in the jmp struct)
 *   A sequence with 1/3 or more of its elements ACGTU is assumed to be DNA
 *   Output is in the file "align.out"
 *
 * The program may create a tmp file in /tmp to hold info about traceback.
 * Original version developed under BSD 4.3 on a vax 8650
 */
#include "nw.h"
#include "day.h"

static  _dbval[26] = {
        1,14,2,13,0,0,4,11,0,0,12,0,3,15,0,0,0,5,6,8,8,7,9,0,10,0
};

static  _pbval[26] = {
        1, 2|(1<<('D'-'A'))|(1<<('N'-'A')), 4, 8, 16, 32, 64,
        128, 256, 0xFFFFFFFF, 1<<10, 1<<11, 1<<12, 1<<13, 1<<14,
        1<<15, 1<<16, 1<<17, 1<<18, 1<<19, 1<<20, 1<<21, 1<<22,
        1<<23, 1<<24, 1<<25|(1<<('E'-'A'))|(1<<('Q'-'A'))
};

main(ac, av)                                                    main
        int     ac;
        char    *av[];
{
        prog = av[0];
        if (ac != 3) {
                fprintf(stderr,"usage: %s file1 file2\n", prog);
                fprintf(stderr,"where file1 and file2 are two dna or two protein sequences.\n");
```

```
            fprintf(stderr,"The sequences can be in upper- or lower-case\n");
            fprintf(stderr,"Any lines beginning with ';' or '<' are ignored\n");
            fprintf(stderr,"Output is in the file \"align.out\"\n");
            exit(1);
        }
        namex[0] = av[1];
        namex[1] = av[2];
        seqx[0] = getseq(namex[0], &len0);
        seqx[1] = getseq(namex[1], &len1);
        xbm = (dna)? _dbval : _pbval;

        endgaps = 0;                    /* 1 to penalize endgaps */
        ofile = "align.out";            /* output file */

        nw();           /* fill in the matrix, get the possible jmps */
        readjmps();     /* get the actual jmps */
        print();        /* print stats, alignment */

        cleanup(0);     /* unlink any tmp files */
}
```

Table 1 (cont.)

```
/* do the alignment, return best score: main()
 * dna: values in Fitch and Smith, PNAS, 80, 1382-1386, 1983
 * pro: PAM 250 values
 * When scores are equal, we prefer mismatches to any gap, prefer
 * a new gap to extending an ongoing gap, and prefer a gap in seqx
 * to a gap in seq y.
 */
nw()                                                                    nw
{
        char            *px, *py;       /* seqs and ptrs */
        int             *ndely, *dely;  /* keep track of dely */
        int             ndelx, delx;    /* keep track of delx */
        int             *tmp;           /* for swapping row0, row1 */
        int             mis;            /* score for each type */
        int             ins0, ins1;     /* insertion penalties */
        register        id;             /* diagonal index */
        register        ij;             /* jmp index */
        register        *col0, *col1;   /* score for curr, last row */
        register        xx, yy;         /* index into seqs */

        dx = (struct diag *)g_calloc("to get diags", len0+len1+1, sizeof(struct diag));

        ndely = (int *)g_calloc("to get ndely", len1+1, sizeof(int));
        dely = (int *)g_calloc("to get dely", len1+1, sizeof(int));
        col0 = (int *)g_calloc("to get col0", len1+1, sizeof(int));
        col1 = (int *)g_calloc("to get col1", len1+1, sizeof(int));
        ins0 = (dna)? DINS0 : PINS0;
        ins1 = (dna)? DINS1 : PINS1;

        smax = -10000;
        if (endgaps) {
                for (col0[0] = dely[0] = -ins0, yy = 1; yy <= len1; yy++) {
                        col0[yy] = dely[yy] = col0[yy-1] - ins1;
                        ndely[yy] = yy;
                }
```

```
        col0[0] = 0;        /* Waterman Bull Math Biol 84 */
    }
    else
        for (yy = 1; yy <= len1; yy++)
            dely[yy] = -ins0;

    /* fill in match matrix
     */
    for (px = seqx[0], xx = 1; xx <= len0; px++, xx++) {
        /* initialize first entry in col
         */
        if (endgaps) {
            if (xx == 1)
                col1[0] = delx = -(ins0+ins1);
            else
                col1[0] = delx = col0[0] - ins1;
            ndelx = xx;
        }
        else {
            col1[0] = 0;
            delx = -ins0;
            ndelx = 0;
        }
```

Table 1 (cont.)

...nw

```
for (py = seqx[1], yy = 1; yy <= len1; py++, yy++) {
        mis = col0[yy-1];
        if (dna)
                mis += (xbm[*px-'A']&xbm[*py-'A'])? DMAT : DMIS;
        else
                mis += _day[*px-'A'][*py-'A'];

        /* update penalty for del in x seq;
         * favor new del over ongong del
         * ignore MAXGAP if weighting endgaps
         */
        if (endgaps || ndely[yy] < MAXGAP) {
                if (col0[yy] - ins0 >= dely[yy]) {
                        dely[yy] = col0[yy] - (ins0+ins1);
                        ndely[yy] = 1;
                } else {
                        dely[yy] -= ins1;
                        ndely[yy]++;
                }
        } else {
                if (col0[yy] - (ins0+ins1) >= dely[yy]) {
                        dely[yy] = col0[yy] - (ins0+ins1);
                        ndely[yy] = 1;
                } else
                        ndely[yy]++;
        }

        /* update penalty for del in y seq;
         * favor new del over ongong del
         */
        if (endgaps || ndelx < MAXGAP) {
                if (col1[yy-1] - ins0 >= delx) {
                        delx = col1[yy-1] - (ins0+ins1);
                        ndelx = 1;
```

```
                } else {
                        delx -= ins1;
                        ndelx++;
                }
        } else {
                if (col1[yy-1] - (ins0+ins1) >= delx) {
                        delx = col1[yy-1] - (ins0+ins1);
                        ndelx = 1;
                } else
                        ndelx++;
        }

        /* pick the maximum score; we're favoring
         * mis over any del and delx over dely
         */
```

Table 1 (cont.)

...nw

```
id = xx - yy + len1 - 1;
if (mis >= delx && mis >= dely[yy])
        col1[yy] = mis;
else if (delx >= dely[yy]) {
        col1[yy] = delx;
        ij = dx[id].ijmp;
        if (dx[id].jp.n[0] && (!dna || (ndelx >= MAXJMP
        && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINSO)) {
                dx[id].ijmp++;
                if (++ij >= MAXJMP) {
                        writejmps(id);
                        ij = dx[id].ijmp = 0;
                        dx[id].offset = offset;
                        offset += sizeof(struct jmp) + sizeof(offset);
                }
        }
        dx[id].jp.n[ij] = ndelx;
        dx[id].jp.x[ij] = xx;
        dx[id].score = delx;
}
else {
        col1[yy] = dely[yy];
        ij = dx[id].ijmp;
if (dx[id].jp.n[0] && (!dna || (ndely[yy] >= MAXJMP
        && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINSO)) {
                dx[id].ijmp++;
                if (++ij >= MAXJMP) {
                        writejmps(id);
                        ij = dx[id].ijmp = 0;
                        dx[id].offset = offset;
                        offset += sizeof(struct jmp) + sizeof(offset);
                }
        }
        dx[id].jp.n[ij] = -ndely[yy];
```

```
                        dx[id].jp.x[ij] = xx;
                        dx[id].score = dely[yy];
                    }
                    if (xx == len0 && yy < len1) {
                        /* last col
                         */
                        if (endgaps)
                            col1[yy] -= ins0+ins1*(len1-yy);
                        if (col1[yy] > smax) {
                            smax = col1[yy];
                            dmax = id;
                        }
                    }
                }
                if (endgaps && xx < len0)
                    col1[yy-1] -= ins0+ins1*(len0-xx);
                if (col1[yy-1] > smax) {
                    smax = col1[yy-1];
                    dmax = id;
                }
                tmp = col0; col0 = col1; col1 = tmp;
            }
            (void) free((char *)ndely);
            (void) free((char *)dely);
            (void) free((char *)col0);
            (void) free((char *)col1);                    }
```

Table 1 (cont.)

```
/*
 *
 * print() -- only routine visible outside this module
 *
 * static:
 * getmat() -- trace back best path, count matches: print()
 * pr_align() -- print alignment of described in array p[]: print()
 * dumpblock() -- dump a block of lines with numbers, stars: pr_align()
 * nums() -- put out a number line: dumpblock()
 * putline() -- put out a line (name, [num], seq, [num]): dumpblock()
 * stars() - -put a line of stars: dumpblock()
 * stripname() -- strip any path and prefix from a seqname
 */

#include "nw.h"

#define SPC     3
#define P_LINE  256     /* maximum output line */
#define P_SPC   3       /* space between name or num and seq */

extern  _day[26][26];
int     olen;           /* set output line length */
FILE    *fx;            /* output file */

print()                                                          print
{
        int     lx, ly, firstgap, lastgap;      /* overlap */

        if ((fx = fopen(ofile, "w")) == 0) {
                fprintf(stderr,"%s: can't write %s\n", prog, ofile);
                cleanup(1);
        }
        fprintf(fx, " < first sequence: %s (length = %d)\n", namex[0], len0);
        fprintf(fx, " < second sequence: %s (length = %d)\n", namex[1], len1);
        olen = 60;
```

```
        lx = len0;
        ly = len1;
        firstgap = lastgap = 0;
        if (dmax < len1 - 1) {        /* leading gap in x */
                pp[0].spc = firstgap = len1 - dmax - 1;
                ly -= pp[0].spc;
        }
        else if (dmax > len1 - 1) { /* leading gap in y */
                pp[1].spc = firstgap = dmax - (len1 - 1);
                lx -= pp[1].spc;
        }
        if (dmax0 < len0 - 1) {       /* trailing gap in x */
                lastgap = len0 - dmax0 -1;
                lx -= lastgap;
        }
        else if (dmax0 > len0 - 1) {          /* trailing gap in y */
                lastgap = dmax0 - (len0 - 1);
                ly -= lastgap;
        }
        getmat(lx, ly, firstgap, lastgap);
        pr_align();
}
```

Table 1 (cont.)

```
/*
 * trace back the best path, count matches
 */
static
getmat(lx, ly, firstgap, lastgap)                                    getmat
        int     lx, ly;                 /* "core" (minus endgaps) */
        int     firstgap, lastgap;      /* leading trailing overlap */
{
        int             nm, i0, i1, siz0, siz1;
        char            outx[32];
        double          pct;
        register        n0, n1;
        register char   *p0, *p1;

        /* get total matches, score
         */
        i0 = i1 = siz0 = siz1 = 0;
        p0 = seqx[0] + pp[1].spc;
        p1 = seqx[1] + pp[0].spc;
        n0 = pp[1].spc + 1;
        n1 = pp[0].spc + 1;

        nm = 0;
        while ( *p0 && *p1 ) {
                if (siz0) {
                        p1++;
                        n1++;
                        siz0--;
                }
                else if (siz1) {
                        p0++;
                        n0++;
                        siz1--;
                }
                else {
```

```
                if (xbm[*p0-'A']&xbm[*p1-'A'])
                        nm++;
                if (n0++ == pp[0].x[i0])
                        siz0 = pp[0].n[i0++];
                if (n1++ == pp[1].x[i1])
                        siz1 = pp[1].n[i1++];
                p0++;
                p1++;
        }
}

/* pct homology:
 * if penalizing endgaps, base is the shorter seq
 * else, knock off overhangs and take shorter core
 */
if (endgaps)
        lx = (len0 < len1)? len0 : len1;
else
        lx = (lx < ly)? lx : ly;
pct = 100.*(double)nm/(double)lx;
fprintf(fx, "\n");
fprintf(fx, " < %d match%s in an overlap of %d: %.2f percent similarity\n",
        nm, (nm == 1)? "" : "es", lx, pct);
```

Table 1 (cont.)

```
        fprintf(fx, " < gaps in first sequence: %d", gapx);                    ...getmat
        if (gapx) {
            (void) sprintf(outx, " (%d %s%s)",
                ngapx, (dna)? "base":"residue", (ngapx == 1)? "":"s");
            fprintf(fx,"%s", outx);

        fprintf(fx, ", gaps in second sequence: %d", gapy);
        if (gapy) {
            (void) sprintf(outx, " (%d %s%s)",
                ngapy, (dna)? "base":"residue", (ngapy == 1)? "":"s");
            fprintf(fx,"%s", outx);
        }
        if (dna)
            fprintf(fx,
            "\n< score: %d (match = %d, mismatch = %d, gap penalty = %d + %d per base)\n",
            smax, DMAT, DMIS, DINS0, DINS1);
        else
            fprintf(fx,
            "\n< score: %d (Dayhoff PAM 250 matrix, gap penalty = %d + %d per residue)\n",
            smax, PINS0, PINS1);
        if (endgaps)
            fprintf(fx,
            " < endgaps penalized. left endgap: %d %s%s, right endgap: %d %s%s\n",
            firstgap, (dna)? "base" : "residue", (firstgap == 1)? "" : "s",
            lastgap, (dna)? "base" : "residue", (lastgap == 1)? "" : "s");
        else
            fprintf(fx, " < endgaps not penalized\n");
    }

    static      nm;         /* matches in core -- for checking */
    static      lmax;       /* lengths of stripped file names */
    static      ij[2];      /* jmp index for a path */
    static      nc[2];      /* number at start of current line */
    static      ni[2];      /* current elem number -- for gapping */
```

```
static          siz[2];
static char     *ps[2];         /* ptr to current element */
static char     *po[2];         /* ptr to next output char slot */
static char     out[2][P_LINE]; /* output line */
static char     star[P_LINE];   /* set by stars() */

/*
 * print alignment of described in struct path pp[]
 */
static
pr_align()                                                    pr_align
{
        int             nn;     /* char count */
        int             more;
        register        i;

        for (i = 0, lmax = 0; i < 2; i++) {
                nn = stripname(namex[i]);
                if (nn > lmax)
                        lmax = nn;

                nc[i] = 1;
                ni[i] = 1;
                siz[i] = ij[i] = 0;
                ps[i] = seqx[i];
                po[i] = out[i];                 }
```

Table 1 (cont.)

```
for (nn = nm = 0, more = 1; more; ) {                                    ...pr_align
        for (i = more = 0; i < 2; i++) {
                /*
                 * do we have more of this sequence?
                 */
                if (!*ps[i])
                        continue;

                more++;

                if (pp[i].spc) {        /* leading space */
                        *po[i]++ = ' ';
                        pp[i].spc--;
                }
                else if (siz[i]) {      /* in a gap */
                        *po[i]++ = '-';
                        siz[i]--;
                }
                else {                  /* we're putting a seq element
                                         */
                        *po[i] = *ps[i];
                        if (islower(*ps[i]))
                                *ps[i] = toupper(*ps[i]);
                        po[i]++;
                        ps[i]++;

                        /*
                         * are we at next gap for this seq?
                         */
                        if (ni[i] == pp[i].x[ij[i]]) {
                                /*
                                 * we need to merge all gaps
                                 * at this location
                                 */
```

```
                                        siz[i] = pp[i].n[ij[i]++];
                                        while (ni[i] -- pp[i].x[ij[i]])
                                                siz[i] += pp[i].n[ij[i]++];
                                }
                                ni[i]++;
                        }
                }
                if (++nn -- olen || !more && nn) {
                        dumpblock();
                        for (i = 0; i < 2; i++)
                                po[i] = out[i];
                        nn = 0;
                }
        }
}

/*
 * dump a block of lines, including numbers, stars: pr_align()
 */
static
dumpblock()                                                    dumpblock
{
        register i;

        for (i = 0; i < 2; i++)
                *po[i]-- = '\0';
```

Table 1 (cont.)

...dumpblock

```
        (void) putc('\n', fx);
        for (i = 0; i < 2; i++) {
                if (*out[i] && (*out[i] != ' ' || *(po[i]) != ' ')) {
                        if (i == 0)
                                nums(i);
                        if (i == 0 && *out[1])
                                stars();
                        putline(i);
                        if (i == 0 && *out[1])
                                fprintf(fx, star);
                        if (i == 1)
                                nums(i);
                }
        }
}

/*
 * put out a number line: dumpblock()
 */
static
nums(ix)                                                        nums
        int     ix;     /* index in out[] holding seq line */
{
        char            nline[P_LINE];
        register        i, j;
        register char   *pn, *px, *py;

        for (pn = nline, i = 0; i < lmax+P_SPC; i++, pn++)
                *pn = ' ';
        for (i = nc[ix], py = out[ix]; *py; py++, pn++) {
                if (*py == ' ' || *py == '-')
                        *pn = ' ';
                else {
```

```
                        if (i%10 == 0 || (i == 1 && nc[ix] != 1)) {
                                j = (i < 0)? -i : i;
                                for (px = pn; j; j /= 10, px--)
                                        *px = j%10 + '0';
                                if (i < 0)
                                        *px = '-';
                        }
                        else
                                *pn = ' ';
                        i++;
                }
        }
        *pn = '\0';
        nc[ix] = i;
        for (pn = nline; *pn; pn++)
                (void) putc(*pn, fx);
        (void) putc('\n', fx);
}

/*
 * put out a line (name, [num], seq, [num]): dumpblock()
 */
static
putline(ix)                                                    putline
        int     ix;                     {
```

Table 1 (cont.)

```
                                                                    ...putline
        int             i;
        register char   *px;

        for (px = namex[ix], i = 0; *px && *px != ':'; px++, i++)
                (void) putc(*px, fx);
        for (; i < lmax+P_SPC; i++)
                (void) putc(' ', fx);

        /* these count from 1:
         * ni[] is current element (from 1)
         * nc[] is number at start of current line
         */
        for (px = out[ix]; *px; px++)
                (void) putc(*px&0x7F, fx);
        (void) putc('\n', fx);
}

/*
 * put a line of stars (seqs always in out[0], out[1]): dumpblock()
 */
static
stars()                                                             stars
{
        int             i;
        register char   *p0, *p1, cx, *px;

        if (!*out[0] || (*out[0] == ' ' && *(p0[0]) == ' ') ||
            !*out[1] || (*out[1] == ' ' && *(p0[1]) == ' '))
                return;
        px = star;
        for (i = lmax+P_SPC; i; i--)
                *px++ = ' ';
```

```
        for (p0 = out[0], p1 = out[1]; *p0 && *p1; p0++, p1++) {
                if (isalpha(*p0) && isalpha(*p1)) {

                        if (xbm[*p0-'A']&xbm[*p1-'A']) {
                                cx = '*';
                                nm++;
                        }
                        else if (!dna && _day[*p0-'A'][*p1-'A'] > 0)
                                cx = '.';
                        else
                                cx = ' ';
                }
                else
                        cx = ' ';
                *px++ = cx;
        }
        *px++ = '\n';
        *px = '\0';
}
```

Table 1 (cont.)

```
/*
 * strip path or prefix from pn, return len: pr_align()
 */
static
stripname(pn)                                                    stripname
        char    *pn;    /* file name (may be path) */
{
        register char   *px, *py;

        py = 0;
        for (px = pn; *px; px++)
                if (*px == '/')
                        py = px + 1;
        if (py)
                (void) strcpy(pn, py);
        return(strlen(pn));

}
```

Table 1 (cont.)

```
/*
 * cleanup() -- cleanup any tmp file
 * getseq() -- read in seq, set dna, len, maxlen
 * g_calloc() -- calloc() with error checkin
 * readjmps() -- get the good jmps, from tmp file if necessary
 * writejmps() -- write a filled array of jmps to a tmp file: nw()
 */
#include "nw.h"
#include <sys/file.h>

char    *jname = "/tmp/homgXXXXXX";             /* tmp file for jmps */
FILE    *fj;

int     cleanup();                      /* cleanup tmp file */
long    lseek();

/*
 * remove any tmp file if we blow
 */
cleanup(i)                                                      cleanup
        int     i;
{
        if (fj)
                (void) unlink(jname);
        exit(i);
}

/*
 * read, return ptr to seq, set dna, len, maxlen
 * skip lines starting with ';', '<', or '>'
 * seq in upper or lower case
 */
char    *
getseq(file, len)                                               getseq
        char    *file;  /* file name */
```

```
        int     *len;   /* seq len */
{
        char            line[1024], *pseq;
        register char   *px, *py;
        int             natgc, tlen;
        FILE            *fp;

        if ((fp = fopen(file,"r")) == 0) {
                fprintf(stderr,"%s: can't read %s\n", prog, file);
                exit(1);
        }
        tlen = natgc = 0;
        while (fgets(line, 1024, fp)) {
                if (*line == ';' || *line == '<' || *line == '>')
                        continue;
                for (px = line; *px != '\n'; px++)
                        if (isupper(*px) || islower(*px))
                                tlen++;
        }
        if ((pseq = malloc((unsigned)(tlen+6))) == 0) {
                fprintf(stderr,"%s: malloc() failed to get %d bytes for %s\n", prog, tlen+6, file);
                exit(1);
        }
        pseq[0] = pseq[1] = pseq[2] = pseq[3] = '\0';
```

Table 1 (cont)

```
                                                                    ...getseq
        py = pseq + 4;
        *len = tlen;
        rewind(fp);

        while (fgets(line, 1024, fp)) {
                if (*line == ';' || *line == '<' || *line == '>')
                        continue;
                for (px = line; *px != '\n'; px++) {
                        if (isupper(*px))
                                *py++ = *px;
                        else if (islower(*px))
                                *py++ = toupper(*px);
                        if (index("ATGCU",*(py-1)))
                                natgc++;
                }
        }
        *py++ = '\0';
        *py = '\0';
        (void) fclose(fp);
        dna = natgc > (tlen/3);
        return(pseq+4);
}

char    *
g_calloc(msg, nx, sz)                                               g_calloc
        char    *msg;           /* program, calling routine */
        int     nx, sz;         /* number and size of elements */
{
        char            *px, *calloc();

        if ((px = calloc((unsigned)nx, (unsigned)sz)) == 0) {
                if (*msg) {
                        fprintf(stderr, "%s: g_calloc() failed %s (n=%d, sz=%d)\n", prog, msg, nx, sz);
                        exit(1);
```

```
                }
        }
        return(px);
}

/*
 * get final jmps from dx[] or tmp file, set pp[], reset dmax: main()
 */
readjmps()                                                              readjmps
{
        int             fd = -1;
        int             siz, i0, i1;
        register i, j, xx;

        if (fj) {
                (void) fclose(fj);
                if ((fd = open(jname, O_RDONLY, 0)) < 0) {
                        fprintf(stderr, "%s: can't open() %s\n", prog, jname);
                        cleanup(1);
                }
        }
        for (i = i0 = i1 = 0, dmax0 = dmax, xx = len0; ; i++) {
                while (1) {
                        for (j = dx[dmax].ijmp; j >= 0 && dx[dmax].jp.x[j] >= xx; j--)
                                ;
```

Table 1 (cont.)

...readjmps

```
                if (j < 0 && dx[dmax].offset && fj) {
                        (void) lseek(fd, dx[dmax].offset, 0);
                        (void) read(fd, (char *)&dx[dmax].jp, sizeof(struct jmp));
                        (void) read(fd, (char *)&dx[dmax].offset, sizeof(dx[dmax].offset));
                        dx[dmax].ijmp = MAXJMP-1;
                }
                else
                        break;
        }
        if (i >= JMPS) {
                fprintf(stderr, "%s: too many gaps in alignment\n", prog);
                cleanup(1);
        }
        if (j >= 0) {
                siz = dx[dmax].jp.n[j];
                xx = dx[dmax].jp.x[j];
                dmax += siz;
                if (siz < 0) {                  /* gap in second seq */
                        pp[1].n[i1] = -siz;
                        xx += siz;
                        /* id = xx - yy + len1 - 1
                         */
                        pp[1].x[i1] = xx - dmax + len1 - 1;
                        gapy++;
                        ngapy -= siz;
/* ignore MAXGAP when doing endgaps */
                        siz = (-siz < MAXGAP || endgaps)? -siz : MAXGAP;
                        i1++;
                }
                else if (siz > 0) {/* gap in first seq */
                        pp[0].n[i0] = siz;
                        pp[0].x[i0] = xx;
                        gapx++;
                        ngapx += siz;
```

```
/* ignore MAXGAP when doing endgaps */
                        siz = (siz < MAXGAP || endgaps)? siz : MAXGAP;
                        i0++;
                }
        }
        else
                break;
}

/* reverse the order of jmps
 */
for (j = 0, i0--; j < i0; j++, i0--) {
        i = pp[0].n[j]; pp[0].n[j] = pp[0].n[i0]; pp[0].n[i0] = i;
        i = pp[0].x[j]; pp[0].x[j] = pp[0].x[i0]; pp[0].x[i0] = i;
}
for (j = 0, i1--; j < i1; j++, i1--) {
        i = pp[1].n[j]; pp[1].n[j] = pp[1].n[i1]; pp[1].n[i1] = i;
        i = pp[1].x[j]; pp[1].x[j] = pp[1].x[i1]; pp[1].x[i1] = i;
}
if (fd >= 0)
        (void) close(fd);
if (fj) {
        (void) unlink(jname);
        fj = 0;
        offset = 0;
}                       }
```

Table 1 (cont.)

```
/*
 * write a filled jmp struct offset of the prev one (if any): nw()
 */
writejmps(ix)                                                    writejmps
        int     ix;
{
        char    *mktemp();

        if (!fj) {
                if (mktemp(jname) < 0) {
                        fprintf(stderr, "%s: can't mktemp() %s\n", prog, jname);
                        cleanup(1);
                }
                if ((fj = fopen(jname, "w")) == 0) {
                        fprintf(stderr, "%s: can't write %s\n", prog, jname);
                        exit(1);
                }
        }
        (void) fwrite((char *)&dx[ix].jp, sizeof(struct jmp), 1, fj);
        (void) fwrite((char *)&dx[ix].offset, sizeof(dx[ix].offset), 1, fj);
}
```

Table 2

PRO	XXXXXXXXXXXXXX	(Length = 15 amino acids)
Comparison Protein	XXXXXYYYYYY	(Length = 12 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 15 = 33.3%

Table 3

PRO	XXXXXXXXXX	(Length = 10 amino acids)
Comparison Protein	XXXXXYYYYYYZZYZ	(Length = 15 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 10 = 50%

Table 4

PRO-DNA	NNNNNNNNNNNNNN	(Length = 14 nucleotides)
Comparison DNA	NNNNNNLLLLLLLLL	(Length = 16 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

6 divided by 14 = 42.9%

Table 5

PRO-DNA	NNNNNNNNNNN	(Length = 12 nucleotides)
Comparison DNA	NNNNLLLVV	(Length = 9 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

4 divided by 12 = 33.3%

II. Compositions and Methods of the Invention

A. Full-length CHEPO Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as CHEPO. In particular, DNA encoding a CHEPO polypeptide has been identified and isolated, as disclosed in further detail in the Examples below.

B. CHEPO Variants

In addition to the full-length native sequence CHEPO polypeptides described herein, it is contemplated that CHEPO variants can be prepared. CHEPO variants can be prepared by introducing appropriate nucleotide changes into the CHEPO DNA, and/or by synthesis of the desired CHEPO polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the CHEPO, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence CHEPO or in various domains of the CHEPO described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the CHEPO that results in a change in the amino acid sequence of the CHEPO as compared with the native sequence CHEPO. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the CHEPO. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the CHEPO with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of about 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity exhibited by the full-length or mature native sequence.

CHEPO polypeptide fragments are provided herein. Such fragments may be truncated at the N-terminus or C-terminus, or may lack internal residues, for example, when compared with a full length native protein. Certain fragments lack amino acid residues that are not essential for a desired biological activity of the CHEPO polypeptide.

CHEPO fragments may be prepared by any of a number of conventional techniques. Desired peptide fragments may be chemically synthesized. An alternative approach involves generating CHEPO fragments by enzymatic digestion, e.g., by treating the protein with an enzyme known to cleave proteins at sites defined by particular amino acid residues, or by digesting the DNA with suitable restriction enzymes and isolating the desired fragment. Yet another suitable technique involves isolating and amplifying a DNA fragment encoding a desired polypeptide fragment, by polymerase chain reaction (PCR). Oligonucleotides that define the desired termini of the DNA fragment are employed at the 5' and 3' primers in the PCR. Preferably, CHEPO polypeptide fragments share at least one biological and/or immunological activity with the native CHEPO polypeptides shown in Figure 3 (SEQ ID NOS:2 and 5).

In particular embodiments, conservative substitutions of interest are shown in Table 6 under the heading of preferred substitutions. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions in Table 6, or as further described below in reference to amino acid classes, are introduced and the products screened.

Table 6

Original Residue	Exemplary Substitutions	Preferred Substitutions
Ala (A)	val; leu; ile	val
Arg (R)	lys; gln; asn	lys
Asn (N)	gln; his; lys; arg	gln
Asp (D)	glu	glu
Cys (C)	ser	ser
Gln (Q)	asn	asn
Glu (E)	asp	asp
Gly (G)	pro; ala	ala
His (H)	asn; gln; lys; arg	arg
Ile (I)	leu; val; met; ala; phe; norleucine	leu
Leu (L)	norleucine; ile; val; met; ala; phe	ile
Lys (K)	arg; gln; asn	arg
Met (M)	leu; phe; ile	leu
Phe (F)	leu; val; ile; ala; tyr	leu
Pro (P)	ala	ala
Ser (S)	thr	thr
Thr (T)	ser	ser
Trp (W)	tyr; phe	tyr
Tyr (Y)	trp; phe; thr; ser	phe
Val (V)	ile; leu; met; phe; ala; norleucine	leu

Substantial modifications in function or immunological identity of the CHEPO polypeptide are accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the

molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues are divided into groups based on common side-chain properties:

(1) hydrophobic: norleucine, met, ala, val, leu, ile;

(2) neutral hydrophilic: cys, ser, thr;

(3) acidic: asp, glu;

(4) basic: asn, gln, his, lys, arg;

(5) residues that influence chain orientation: gly, pro; and

(6) aromatic: trp, tyr, phe.

Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Such substituted residues also may be introduced into the conservative substitution sites or, more preferably, into the remaining (non-conserved) sites.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the CHEPO variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant [Cunningham and Wells, Science, 244: 1081-1085 (1989)]. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

C. Modifications of CHEPO

Covalent modifications of CHEPO are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a CHEPO polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C- terminal residues of the CHEPO. Derivatization with bifunctional agents is useful, for instance, for crosslinking CHEPO to a water-insoluble support matrix or surface for use in the method for purifying anti-CHEPO antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the -amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the CHEPO polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence CHEPO (either by removing the underlying glycosylation site or by deleting the glycosylation by chemical and/or enzymatic means), and/or adding one or more glycosylation sites that are not present in the native sequence CHEPO. In addition, the phrase includes qualitative changes in the glycosylation of the native proteins, involving a change in the nature and proportions of the various carbohydrate moieties present.

Addition of glycosylation sites to the CHEPO polypeptide may be accomplished by altering the amino acid sequence. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence CHEPO (for O-linked glycosylation sites). The CHEPO amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the CHEPO polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the CHEPO polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the CHEPO polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of CHEPO comprises linking the CHEPO polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol (PEG), polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

The CHEPO of the present invention may also be modified in a way to form a chimeric molecule comprising CHEPO fused to another, heterologous polypeptide or amino acid sequence.

In one embodiment, such a chimeric molecule comprises a fusion of the CHEPO with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the CHEPO. The presence of such epitope-tagged forms of the CHEPO can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the CHEPO to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. Various tag

polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an -tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

In an alternative embodiment, the chimeric molecule may comprise a fusion of the CHEPO with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule (also referred to as an immunoadhesin), such a fusion could be to the Fc region of an IgG molecule. The Ig fusions preferably include the substitution of a soluble (transmembrane domain deleted or inactivated) form of a CHEPO polypeptide in place of at least one variable region within an Ig molecule. In a particularly preferred embodiment, the immunoglobulin fusion includes the hinge, CH2 and CH3, or the hinge, CH1, CH2 and CH3 regions of an IgG1 molecule. For the production of immunoglobulin fusions see also US Patent No. 5,428,130 issued June 27, 1995.

D. Preparation of CHEPO

The description below relates primarily to production of CHEPO by culturing cells transformed or transfected with a vector containing CHEPO nucleic acid. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare CHEPO. For instance, the CHEPO sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of the CHEPO may be chemically synthesized separately and combined using chemical or enzymatic methods to produce the full-length CHEPO.

1. Isolation of DNA Encoding CHEPO

DNA encoding CHEPO may be obtained from a cDNA library prepared from tissue believed to possess the CHEPO mRNA and to express it at a detectable level. Accordingly, human CHEPO DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The CHEPO-encoding gene may also be obtained from a genomic library or by known synthetic procedures (e.g., automated nucleic acid synthesis).

Libraries can be screened with probes (such as antibodies to the CHEPO or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate

the gene encoding CHEPO is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined using methods known in the art and as described herein.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for CHEPO production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of eukaryotic cell transfection and prokaryotic cell transformation are known to the ordinarily skilled artisan, for example, $CaCl_2$, $CaPO_4$, liposome-mediated and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transfections have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine,

may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635). Other suitable prokaryotic host cells include Enterobacteriaceae such as *Escherichia, e.g., E. coli, Enterobacter, Erwinia, Klebsiella, Proteus, Salmonella, e.g., Salmonella typhimurium, Serratia, e.g., Serratia marcescans*, and *Shigella*, as well as *Bacilli* such as *B. subtilis* and *B. licheniformis* (*e.g., B. licheniformis* 41P disclosed in DD 266,710 published 12 April 1989), *Pseudomonas* such as *P. aeruginosa*, and *Streptomyces*. These examples are illustrative rather than limiting. Strain W3110 is one particularly preferred host or parent host because it is a common host strain for recombinant DNA product fermentations. Preferably, the host cell secretes minimal amounts of proteolytic enzymes. For example, strain W3110 may be modified to effect a genetic mutation in the genes encoding proteins endogenous to the host, with examples of such hosts including *E. coli* W3110 strain 1A2, which has the complete genotype *tonA* ; *E. coli* W3110 strain 9E4, which has the complete genotype *tonA ptr3*; *E. coli* W3110 strain 27C7 (ATCC 55,244), which has the complete genotype *tonA ptr3 phoA E15 (argF-lac)169 degP ompT kan*r; *E. coli* W3110 strain 37D6, which has the complete genotype *tonA ptr3 phoA E15 (argF-lac)169 degP ompT rbs7 ilvG kan*r; *E. coli* W3110 strain 40B4, which is strain 37D6 with a non-kanamycin resistant *degP* deletion mutation; and an *E. coli* strain having mutant periplasmic protease disclosed in U.S. Patent No. 4,946,783 issued 7 August 1990. Alternatively, *in vitro* methods of cloning, *e.g.*, PCR or other nucleic acid polymerase reactions, are suitable.

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for CHEPO-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism. Others include *Schizosaccharomyces pombe* (Beach and Nurse, Nature, 290: 140 [1981]; EP 139,383 published 2 May 1985); *Kluyveromyces* hosts (U.S. Patent No. 4,943,529; Fleer *et al.*, Bio/Technology, 9: 968-975 (1991)) such as, *e.g., K. lactis* (MW98-8C, CBS683, CBS4574; Louvencourt *et al.*, J. Bacteriol., 737 [1983]), *K. fragilis* (ATCC 12,424), *K. bulgaricus* (ATCC 16,045), *K. wickeramii* (ATCC 24,178), *K. waltii* (ATCC 56,500), *K. drosophilarum* (ATCC 36,906; Van den Berg *et al.*, Bio/Technology, 8: 135 (1990)), *K. thermotolerans*, and *K. marxianus; yarrowia* (EP 402,226); *Pichia pastoris* (EP 183,070; Sreekrishna *et al.*, J. Basic Microbiol., 28: 265-278 [1988]); *Candida; Trichoderma reesia* (EP 244,234); *Neurospora crassa* (Case *et al.*, Proc. Natl. Acad. Sci. USA, 76: 5259-5263 [1979]); *Schwanniomyces* such as *Schwanniomyces occidentalis* (EP 394,538 published 31 October 1990); and filamentous fungi such as, *e.g., Neurospora, Penicillium, Tolypocladium* (WO 91/00357 published 10 January 1991), and *Aspergillus* hosts such as *A. nidulans* (Ballance *et al.*, Biochem. Biophys. Res. Commun., 112: 284-289 [1983]; Tilburn *et al.*, Gene, 26: 205-221 [1983]; Yelton *et al.*, Proc. Natl. Acad. Sci. USA, 81: 1470-1474 [1984]) and *A. niger* (Kelly and Hynes, EMBO J., 4: 475-479 [1985]). Methylotropic yeasts are suitable herein and include, but are not limited to, yeast capable of growth on methanol selected from the genera consisting of *Hansenula, Candida, Kloeckera, Pichia, Saccharomyces, Torulopsis*, and *Rhodotorula*. A list

of specific species that are exemplary of this class of yeasts may be found in C. Anthony, The Biochemistry of Methylotrophs, 269 (1982).

Suitable host cells for the expression of glycosylated CHEPO are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding CHEPO may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The CHEPO may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the CHEPO-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* -factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the CHEPO-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the CHEPO-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the -lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding CHEPO.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

CHEPO transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding the CHEPO by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase,

albumin, -fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the CHEPO coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding CHEPO.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of CHEPO in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence CHEPO polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to CHEPO DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of CHEPO may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of CHEPO can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify CHEPO from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse

phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the CHEPO. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular CHEPO produced.

E. Uses for CHEPO

Nucleotide sequences (or their complement) encoding CHEPO have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. CHEPO nucleic acid will also be useful for the preparation of CHEPO polypeptides by the recombinant techniques described herein.

The full-length native sequence CHEPO cDNA (SEQ ID NO:3), or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length CHEPO cDNA or to isolate still other cDNAs (for instance, those encoding naturally-occurring variants of CHEPO or CHEPO from other species) which have a desired sequence identity to the CHEPO sequence disclosed in Figure 2 (SEQ ID NO:3). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from at least partially novel regions of the nucleotide sequence of SEQ ID NO:3 wherein those regions may be determined without undue experimentation or from genomic sequences including promoters, enhancer elements and introns of native sequence CHEPO. By way of example, a screening method will comprise isolating the coding region of the CHEPO gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the CHEPO gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

Any EST sequences disclosed in the present application may similarly be employed as probes, using the methods disclosed herein.

Other useful fragments of the CHEPO nucleic acids include antisense or sense oligonucleotides comprising a singe-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target CHEPO mRNA (sense) or CHEPO DNA (antisense) sequences. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment of the coding region of CHEPO DNA. Such a fragment generally comprises at least about 14 nucleotides, preferably from about 14 to 30 nucleotides. The ability to derive an antisense or a sense oligonucleotide, based upon a cDNA sequence encoding a given protein is described in, for example, Stein and Cohen (*Cancer Res.* 48:2659, 1988) and van der Krol et al. (*BioTechniques* 6:958, 1988).

Binding of antisense or sense oligonucleotides to target nucleic acid sequences results in the formation of duplexes that block transcription or translation of the target sequence by one of several means, including enhanced degradation of the duplexes, premature termination of transcription or translation, or by other means. The antisense oligonucleotides thus may be used to block expression of CHEPO proteins. Antisense or sense oligonucleotides further comprise oligonucleotides having modified sugar-phosphodiester backbones (or other sugar linkages, such as those described in WO 91/06629) and wherein such sugar linkages are resistant to endogenous nucleases. Such oligonucleotides with resistant sugar linkages are stable *in vivo* (i.e., capable of resisting enzymatic degradation) but retain sequence specificity to be able to bind to target nucleotide sequences.

Other examples of sense or antisense oligonucleotides include those oligonucleotides which are covalently linked to organic moieties, such as those described in WO 90/10048, and other moieties that increases affinity of the oligonucleotide for a target nucleic acid sequence, such as poly-(L-lysine). Further still, intercalating agents, such as ellipticine, and alkylating agents or metal complexes may be attached to sense or antisense oligonucleotides to modify binding specificities of the antisense or sense oligonucleotide for the target nucleotide sequence.

Antisense or sense oligonucleotides may be introduced into a cell containing the target nucleic acid sequence by any gene transfer method, including, for example, $CaPO_4$-mediated DNA transfection, electroporation, or by using gene transfer vectors such as Epstein-Barr virus. In a preferred procedure, an antisense or sense oligonucleotide is inserted into a suitable retroviral vector. A cell containing the target nucleic acid sequence is contacted with the recombinant retroviral vector, either *in vivo* or *ex vivo*. Suitable retroviral vectors include, but are not limited to, those derived from the murine retrovirus M-MuLV, N2 (a retrovirus derived from M-MuLV), or the double copy vectors designated DCT5A, DCT5B and DCT5C (see WO 90/13641).

Sense or antisense oligonucleotides also may be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors. Preferably, conjugation of the ligand binding molecule does not substantially interfere with the ability of the ligand binding molecule to bind to its corresponding molecule or receptor, or block entry of the sense or antisense oligonucleotide or its conjugated version into the cell.

Alternatively, a sense or an antisense oligonucleotide may be introduced into a cell containing the target nucleic acid sequence by formation of an oligonucleotide-lipid complex, as described in WO 90/10448. The sense or antisense oligonucleotide-lipid complex is preferably dissociated within the cell by an endogenous lipase.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related CHEPO coding sequences.

Nucleotide sequences encoding a CHEPO can also be used to construct hybridization probes for mapping the gene which encodes that CHEPO and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for CHEPO encode a protein which binds to another protein (example, where the CHEPO is a receptor), the CHEPO can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor CHEPO can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native CHEPO or a receptor for CHEPO. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode CHEPO or its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding CHEPO can be used to clone genomic DNA encoding CHEPO in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding CHEPO. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for CHEPO transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding CHEPO introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding CHEPO. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of CHEPO can be used to construct a CHEPO "knock out" animal which has a defective or altered gene encoding CHEPO as a result of homologous recombination between the endogenous gene encoding CHEPO and altered genomic DNA encoding CHEPO introduced into an embryonic stem cell of the animal. For example, cDNA encoding CHEPO can be used to clone genomic DNA encoding CHEPO in accordance with established techniques. A portion of the genomic DNA encoding CHEPO can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical*

Approach, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the CHEPO polypeptide.

Nucleic acid encoding the CHEPO polypeptides may also be used in gene therapy. In gene therapy applications, genes are introduced into cells in order to achieve *in vivo* synthesis of a therapeutically effective genetic product, for example for replacement of a defective gene. "Gene therapy" includes both conventional gene therapy where a lasting effect is achieved by a single treatment, and the administration of gene therapeutic agents, which involves the one time or repeated administration of a therapeutically effective DNA or mRNA. Antisense RNAs and DNAs can be used as therapeutic agents for blocking the expression of certain genes *in vivo*. It has already been shown that short antisense oligonucleotides can be imported into cells where they act as inhibitors, despite their low intracellular concentrations caused by their restricted uptake by the cell membrane. (Zamecnik *et al.*, Proc. Natl. Acad. Sci. USA 83, 4143-4146 [1986]). The oligonucleotides can be modified to enhance their uptake, e.g. by substituting their negatively charged phosphodiester groups by uncharged groups.

There are a variety of techniques available for introducing nucleic acids into viable cells. The techniques vary depending upon whether the nucleic acid is transferred into cultured cells *in vitro*, or *in vivo* in the cells of the intended host. Techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, the calcium phosphate precipitation method, etc. The currently preferred *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors and viral coat protein-liposome mediated transfection (Dzau *et al.*, Trends in Biotechnology 11, 205-210 [1993]). In some situations it is desirable to provide the nucleic acid source with an agent that targets the target cells, such as an antibody specific for a cell surface membrane protein or the target cell, a ligand for a receptor on the target cell, etc. Where liposomes are employed, proteins which bind to a cell surface membrane protein associated with endocytosis may be used for targeting and/or to facilitate uptake, e.g. capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half-life. The technique of receptor-mediated endocytosis is described, for example, by Wu *et al.*, J. Biol. Chem. 262, 4429-4432 (1987); and Wagner *et al.*, Proc. Natl. Acad. Sci. USA 87, 3410-3414 (1990). For review of gene marking and gene therapy protocols see Anderson *et al.*, Science 256, 808-813 (1992).

The CHEPO polypeptides described herein may also be employed as molecular weight markers for protein electrophoresis purposes.

The nucleic acid molecules encoding the CHEPO polypeptides or fragments thereof described herein are useful for chromosome identification. In this regard, there exists an ongoing need to identify new chromosome markers, since relatively few chromosome marking reagents, based upon actual sequence data are presently available. Each CHEPO nucleic acid molecule of the present invention can be used as a chromosome marker.

The CHEPO polypeptides and nucleic acid molecules of the present invention may also be used for tissue typing, wherein the CHEPO polypeptides of the present invention may be differentially expressed in one tissue as compared to another. CHEPO nucleic acid molecules will find use for generating probes for PCR, Northern analysis, Southern analysis and Western analysis.

The CHEPO polypeptides described herein may also be employed as therapeutic agents. The CHEPO polypeptides of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the CHEPO product hereof is combined in admixture with a pharmaceutically acceptable carrier vehicle. Therapeutic formulations are prepared for storage by mixing the active ingredient having the desired degree of purity with optional physiologically acceptable carriers, excipients or stabilizers (Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980)), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone, amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN™, PLURONICS™ or PEG.

The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes, prior to or following lyophilization and reconstitution.

Therapeutic compositions herein generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

The route of administration is in accord with known methods, e.g. injection or infusion by intravenous, intraperitoneal, intracerebral, intramuscular, intraocular, intraarterial or intralesional routes, topical administration, or by sustained release systems.

Dosages and desired drug concentrations of pharmaceutical compositions of the present invention may vary depending on the particular use envisioned. The determination of the appropriate dosage or route of administration is well within the skill of an ordinary physician. Animal experiments provide reliable guidance for the determination of effective doses for human therapy. Interspecies scaling of effective doses can be performed following the principles laid down by Mordenti, J. and Chappell, W. "The use of interspecies scaling in toxicokinetics" In Toxicokinetics and New Drug Development, Yacobi *et al.*, Eds., Pergamon Press, New York 1989, pp. 42-96.

When *in vivo* administration of a CHEPO polypeptide or agonist or antagonist thereof is employed, normal dosage amounts may vary from about 10 ng/kg to up to 100 mg/kg of mammal body weight or more per day, preferably about 1 g/kg/day to 10 mg/kg/day, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature; see, for example, U.S. Pat. Nos. 4,657,760; 5,206,344; or 5,225,212. It is anticipated that different formulations will be effective for different treatment compounds and different disorders, that

administration targeting one organ or tissue, for example, may necessitate delivery in a manner different from that to another organ or tissue.

Where sustained-release administration of a CHEPO polypeptide is desired in a formulation with release characteristics suitable for the treatment of any disease or disorder requiring administration of the CHEPO polypeptide, microencapsulation of the CHEPO polypeptide is contemplated. Microencapsulation of recombinant proteins for sustained release has been successfully performed with human growth hormone (rhGH), interferon- (rhIFN-), interleukin-2, and MN rgp120. Johnson *et al.*, Nat. Med., 2: 795-799 (1996); Yasuda, Biomed. Ther., 27: 1221-1223 (1993); Hora *et al.*, Bio/Technology, 8: 755-758 (1990); Cleland, "Design and Production of Single Immunization Vaccines Using Polylactide Polyglycolide Microsphere Systems," in Vaccine Design: The Subunit and Adjuvant Approach, Powell and Newman, eds, (Plenum Press: New York, 1995), pp. 439-462; WO 97/03692, WO 96/40072, WO 96/07399; and U.S Pat. No. 5,654,010.

The sustained-release formulations of these proteins were developed using poly-lactic-coglycolic acid (PLGA) polymer due to its biocompatibility and wide range of biodegradable properties. The degradation products of PLGA, lactic and glycolic acids, can be cleared quickly within the human body. Moreover, the degradability of this polymer can be adjusted from months to years depending on its molecular weight and composition. Lewis, "Controlled release of bioactive agents from lactide/glycolide polymer," in: M. Chasin and R. Langer (Eds.), Biodegradable Polymers as Drug Delivery Systems (Marcel Dekker: New York, 1990), pp. 1-41.

This invention encompasses methods of screening compounds to identify those that mimic the CHEPO polypeptide (agonists) or prevent the effect of the CHEPO polypeptide (antagonists). Screening assays for antagonist drug candidates are designed to identify compounds that bind or complex with the CHEPO polypeptides encoded by the genes identified herein, or otherwise interfere with the interaction of the encoded polypeptides with other cellular proteins. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates.

The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays, and cell-based assays, which are well characterized in the art. All assays for antagonists are common in that they call for contacting the drug candidate with a CHEPO polypeptide encoded by a nucleic acid identified herein under conditions and for a time sufficient to allow these two components to interact.

In binding assays, the interaction is binding and the complex formed can be isolated or detected in the reaction mixture. In a particular embodiment, the CHEPO polypeptide encoded by the gene identified herein or the drug candidate is immobilized on a solid phase, *e.g.*, on a microtiter plate, by covalent or non-covalent attachments. Non-covalent attachment generally is accomplished by coating the solid surface with a solution of the CHEPO polypeptide and drying. Alternatively, an immobilized antibody, *e.g.*, a monoclonal antibody, specific for the CHEPO polypeptide to be immobilized can be used to anchor it to a solid surface. The assay is performed by adding the non-immobilized component, which may be labeled by a detectable label, to the immobilized component, *e.g.*, the coated surface containing the anchored component. When the reaction is complete, the non-reacted components are removed, *e.g.*, by washing, and complexes anchored on the solid surface are detected. When the originally non-immobilized component carries a detectable label, the detection of label immobilized on the surface indicates that complexing occurred. Where the originally non-immobilized

component does not carry a label, complexing can be detected, for example, by using a labeled antibody specifically binding the immobilized complex.

If the candidate compound interacts with but does not bind to a particular CHEPO polypeptide encoded by a gene identified herein, its interaction with that polypeptide can be assayed by methods well known for detecting protein-protein interactions. Such assays include traditional approaches, such as, *e.g.*, cross-linking, co-immunoprecipitation, and co-purification through gradients or chromatographic columns. In addition, protein-protein interactions can be monitored by using a yeast-based genetic system described by Fields and co-workers (Fields and Song, Nature (London), 340: 245-246 (1989); Chien *et al.*, Proc. Natl. Acad. Sci. USA, 88: 9578-9582 (1991)) as disclosed by Chevray and Nathans, Proc. Natl. Acad. Sci. USA, 89: 5789-5793 (1991). Many transcriptional activators, such as yeast GAL4, consist of two physically discrete modular domains, one acting as the DNA-binding domain, the other one functioning as the transcription-activation domain. The yeast expression system described in the foregoing publications (generally referred to as the "two-hybrid system") takes advantage of this property, and employs two hybrid proteins, one in which the target protein is fused to the DNA-binding domain of GAL4, and another, in which candidate activating proteins are fused to the activation domain. The expression of a GAL1-*lac*Z reporter gene under control of a GAL4-activated promoter depends on reconstitution of GAL4 activity via protein-protein interaction. Colonies containing interacting polypeptides are detected with a chromogenic substrate for -galactosidase. A complete kit (MATCHMAKER™) for identifying protein-protein interactions between two specific proteins using the two-hybrid technique is commercially available from Clontech. This system can also be extended to map protein domains involved in specific protein interactions as well as to pinpoint amino acid residues that are crucial for these interactions.

Compounds that interfere with the interaction of a gene encoding a CHEPO polypeptide identified herein and other intra- or extracellular components can be tested as follows: usually a reaction mixture is prepared containing the product of the gene and the intra- or extracellular component under conditions and for a time allowing for the interaction and binding of the two products. To test the ability of a candidate compound to inhibit binding, the reaction is run in the absence and in the presence of the test compound. In addition, a placebo may be added to a third reaction mixture, to serve as positive control. The binding (complex formation) between the test compound and the intra- or extracellular component present in the mixture is monitored as described hereinabove. The formation of a complex in the control reaction(s) but not in the reaction mixture containing the test compound indicates that the test compound interferes with the interaction of the test compound and its reaction partner.

To assay for antagonists, the CHEPO polypeptide may be added to a cell along with the compound to be screened for a particular activity and the ability of the compound to inhibit the activity of interest in the presence of the CHEPO polypeptide indicates that the compound is an antagonist to the CHEPO polypeptide. Alternatively, antagonists may be detected by combining the CHEPO polypeptide and a potential antagonist with membrane-bound CHEPO polypeptide receptors or recombinant receptors under appropriate conditions for a competitive inhibition assay. The CHEPO polypeptide can be labeled, such as by radioactivity, such that the number of CHEPO polypeptide molecules bound to the receptor can be used to determine the effectiveness of the potential antagonist. The gene encoding the receptor can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Coligan *et al.*, Current

Protocols in Immun., 1(2): Chapter 5 (1991). Preferably, expression cloning is employed wherein polyadenylated RNA is prepared from a cell responsive to the CHEPO polypeptide and a cDNA library created from this RNA is divided into pools and used to transfect COS cells or other cells that are not responsive to the CHEPO polypeptide. Transfected cells that are grown on glass slides are exposed to labeled CHEPO polypeptide. The CHEPO polypeptide can be labeled by a variety of means including iodination or inclusion of a recognition site for a site-specific protein kinase. Following fixation and incubation, the slides are subjected to autoradiographic analysis. Positive pools are identified and sub-pools are prepared and re-transfected using an interactive sub-pooling and re-screening process, eventually yielding a single clone that encodes the putative receptor.

As an alternative approach for receptor identification, labeled CHEPO polypeptide can be photoaffinity-linked with cell membrane or extract preparations that express the receptor molecule. Cross-linked material is resolved by PAGE and exposed to X-ray film. The labeled complex containing the receptor can be excised, resolved into peptide fragments, and subjected to protein micro-sequencing. The amino acid sequence obtained from micro- sequencing would be used to design a set of degenerate oligonucleotide probes to screen a cDNA library to identify the gene encoding the putative receptor.

In another assay for antagonists, mammalian cells or a membrane preparation expressing the receptor would be incubated with labeled CHEPO polypeptide in the presence of the candidate compound. The ability of the compound to enhance or block this interaction could then be measured.

More specific examples of potential antagonists include an oligonucleotide that binds to the fusions of immunoglobulin with CHEPO polypeptide, and, in particular, antibodies including, without limitation, poly- and monoclonal antibodies and antibody fragments, single-chain antibodies, anti-idiotypic antibodies, and chimeric or humanized versions of such antibodies or fragments, as well as human antibodies and antibody fragments. Alternatively, a potential antagonist may be a closely related protein, for example, a mutated form of the CHEPO polypeptide that recognizes the receptor but imparts no effect, thereby competitively inhibiting the action of the CHEPO polypeptide.

Another potential CHEPO polypeptide antagonist is an antisense RNA or DNA construct prepared using antisense technology, where, *e.g.*, an antisense RNA or DNA molecule acts to block directly the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes the mature CHEPO polypeptides herein, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res., 6: 3073 (1979); Cooney *et al.*, Science, 241: 456 (1988); Dervan *et al.*, Science, 251: 1360 (1991)), thereby preventing transcription and the production of the CHEPO polypeptide. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into the CHEPO polypeptide (antisense - Okano, Neurochem., 56: 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression (CRC Press: Boca Raton, FL, 1988). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of the CHEPO polypeptide. When antisense DNA is used,

oligodeoxyribonucleotides derived from the translation-initiation site, *e.g.*, between about -10 and +10 positions of the target gene nucleotide sequence, are preferred.

Potential antagonists include small molecules that bind to the active site, the receptor binding site, or growth factor or other relevant binding site of the CHEPO polypeptide, thereby blocking the normal biological activity of the CHEPO polypeptide. Examples of small molecules include, but are not limited to, small peptides or peptide-like molecules, preferably soluble peptides, and synthetic non-peptidyl organic or inorganic compounds.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization to the complementary target RNA, followed by endonucleolytic cleavage. Specific ribozyme cleavage sites within a potential RNA target can be identified by known techniques. For further details see, *e.g.*, Rossi, Current Biology, 4: 469-471 (1994), and PCT publication No. WO 97/33551 (published September 18, 1997).

Nucleic acid molecules in triple-helix formation used to inhibit transcription should be single-stranded and composed of deoxynucleotides. The base composition of these oligonucleotides is designed such that it promotes triple-helix formation via Hoogsteen base-pairing rules, which generally require sizeable stretches of purines or pyrimidines on one strand of a duplex. For further details see, *e.g.*, PCT publication No. WO 97/33551, *supra*.

These small molecules can be identified by any one or more of the screening assays discussed hereinabove and/or by any other screening techniques well known for those skilled in the art.

F. Anti-CHEPO Antibodies

The present invention further provides anti-CHEPO antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-CHEPO antibodies may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the CHEPO polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-CHEPO antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit

lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the CHEPO polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against CHEPO. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source

of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Human and Humanized Antibodies

The anti-CHEPO antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-

525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)]. Similarly, human antibodies can be made by introducing of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks *et al.*, Bio/Technology 10, 779-783 (1992); Lonberg *et al.*, Nature 368 856-859 (1994); Morrison, Nature 368, 812-13 (1994); Fishwild *et al.*, Nature Biotechnology 14, 845-51 (1996); Neuberger, Nature Biotechnology 14, 826 (1996); Lonberg and Huszar, Intern. Rev. Immunol. 13 65-93 (1995).

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for the CHEPO, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate

expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory cavities of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab)$_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared can be prepared using chemical linkage. Brennan *et al.*, Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab)$_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab -TNB derivatives is then reconverted to the Fab -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Fab fragments may be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby *et al.*, J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab)$_2$ molecule. Each Fab fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various technique for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny *et al.*, J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The diabody technology described by Hollinger *et al.*, Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another

strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber *et al.*, J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt *et al.*, J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies may bind to two different epitopes on a given CHEPO polypeptide herein. Alternatively, an anti-CHEPO polypeptide arm may be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc R), such as Fc RI (CD64), Fc RII (CD32) and Fc RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular CHEPO polypeptide. Bispecific antibodies may also be used to localize cytotoxic agents to cells which express a particular CHEPO polypeptide. These antibodies possess a CHEPO-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the CHEPO polypeptide and further binds tissue factor (TF).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

6. Effector Function Engineering

It may be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) may be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated may have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron *et al.*, J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity may also be prepared using heterobifunctional cross-linkers as described in Wolff *et al.* Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and may thereby have enhanced complement lysis and ADCC capabilities. See Stevenson *et al.*, Anti-Cancer Drug Design, 3: 219-230 (1989).

7. Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta *et al.*, Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody may be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is conjugated to a cytotoxic agent (*e.g.*, a radionucleotide).

8. Immunoliposomes

The antibodies disclosed herein may also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang *et al.*, Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin *et al.*, J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

9. Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a CHEPO polypeptide identified herein, as well as other molecules identified by the screening assays disclosed hereinbefore, can be administered for the treatment of various disorders in the form of pharmaceutical compositions.

If the CHEPO polypeptide is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, lipofections or liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.,* Marasco *et al.,* Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein may also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition may comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients may also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions. Such techniques are disclosed in Remington's Pharmaceutical Sciences, *supra*.

The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations may be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e.g.,* films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated antibodies remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37 C, resulting in a loss of biological activity and possible changes in immunogenicity. Rational strategies can be devised for stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be

intermolecular S-S bond formation through thio-disulfide interchange, stabilization may be achieved by modifying sulfhydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

G. Uses for anti-CHEPO Antibodies

The anti-CHEPO antibodies of the invention have various utilities. For example, anti-CHEPO antibodies may be used in diagnostic assays for CHEPO, *e.g.*, detecting its expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-CHEPO antibodies also are useful for the affinity purification of CHEPO from recombinant cell culture or natural sources. In this process, the antibodies against CHEPO are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the CHEPO to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the CHEPO, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the CHEPO from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Manassas, VA.

EXAMPLE 1

Isolation of Nucleic Acid Encoding CHEPO

Genomic DNA was isolated from two chimpanzee cell lines (ATCC CRL1609 and CRL1857) using a Qiagen kit (cat#10262) as recommended by the manufacturer's instructions. The chimp Epo gene was then obtained on 3 separate fragments by PCR using 1 g of genomic DNA as template and the following primer pairs:

EPO.F: 5'-ACCGCGCCCCCTGGACAG-3' (SEQ ID NO:12)

EPO.INT1.R: 5'-CATCCACTTCTCCGGCCAAACTTCA-3' (SEQ ID NO:13)

EPO.INT1F: 5'-TTTGGCCGGAGAAGTGGATGC-3' (SEQ ID NO:14)

EPO.INT4R: 5'-TCACTCACTCACTCATTCATTCATTCATTCA-3' (SEQ ID NO:15)

EPO.INT4F: 5'-GTTGAATGAATGATTGAATGAATGAGTGA-3' (SEQ ID NO:16)

EPO.R: 5'-GCACTGGAGTGTCCATGGGACAG-3' (SEQ ID NO:17)

Each PCR reaction contained 5 l of 10x PCR Buffer (Perkin Elmer), 1 l dNTP (20mM), 1 g genomic DNA, 1 l of each primer, 1 l of Taq polymerase (Clontech) and H_2O to bring the total volume to 50 l. The reaction was first denatured for 4 min. at 94 C then amplified for 40 cycles of 1 min. at 94 C, 1 min. at 65 C or 66 C then extended for 1 min. at 72 C. A last extension step of 5 min. at 72 C was performed. The reaction was then analyzed on agarose gel. PCR product of 500bp, 1200bp and 750bp were observed for each PCR product, respectively. The PCR reactions were then purified using a Wizard kit (Promega cat # A7170) then directly sequenced. DNA sequencing of the PCR products was done using an Applied Biosystems 377 DNA Sequencer (PE/Applied Biosystems, Foster City, CA). The chemistry used was DYE Terminator Cycle Sequencing with dRhodamine and BIG DYE terminators (PE/Applied Biosystems, Foster City, CA). Sequence assembly and editing done with Sequencher software (Gene Codes, Ann Arbor, MI).

The 5 coding exons were identified by homology with the human erythropoietin sequence and assembled into a predicted full length cDNA. The coding region of CHEPO cDNA is 579 nucleotides long (Figure 1) and encodes a predicted protein of 193 amino acids (Figure 3). There are 3 putative signal cleavage sites predicted after amino acid residue 22, 24 and 27. In accordance with the N-terminus of human Epo, the latest one is likely to correspond to the cleavage site for chimp Epo. The hydrophobic 27 amino acid signal peptide is followed by a 166 amino acid long mature protein containing 3 potential N-glycosylation sites. A single nucleotide polymorphism is present in the predicted sequence obtained from CRL1609 and changes the protein sequence at amino acid position 142 from a Q to a K. Alignment of the chimp Epo protein with the human sequence indicates a single change at amino acid position 84 (Figure 3).

EXAMPLE 2

Use of CHEPO cDNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding CHEPO as a hybridization probe.

DNA comprising the coding sequence of full-length or mature CHEPO (as shown in Figure 2, SEQ ID NO:3) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of CHEPO) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled CHEPO-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence CHEPO can then be identified using standard techniques known in the art.

EXAMPLE 3

Expression of CHEPO in *E. coli*

This example illustrates preparation of an unglycosylated form of CHEPO by recombinant expression in *E. coli*.

The DNA sequence encoding CHEPO (SEQ ID NO:3) is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the CHEPO coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized CHEPO protein can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

CHEPO may be expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding CHEPO is initially amplified using selected PCR primers. The primers will contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences are then ligated into an expression vector, which is used to transform an

E. coli host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(laclq). Transformants are first grown in LB containing 50 mg/ml carbenicillin at 30 C with shaking until an O.D.600 of 3-5 is reached. Cultures are then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g $(NH_4)_2SO_4$, 0.71 g sodium citrate 2H2O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM $MgSO_4$) and grown for approximately 20-30 hours at 30 C with shaking. Samples are removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets are frozen until purification and refolding.

E. coli paste from 0.5 to 1 L fermentations (6-10 g pellets) is resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution is stirred overnight at 4 C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution is centrifuged at 40,000 rpm in a Beckman Ultracentifuge for 30 min. The supernatant is diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. The clarified extract is loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column is washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein is eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein are pooled and stored at 4 C. Protein concentration is estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins are refolded by diluting the sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes are chosen so that the final protein concentration is between 50 to 100 micrograms/ml. The refolding solution is stirred gently at 4 C for 12-36 hours. The refolding reaction is quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution is filtered through a 0.22 micron filter and acetonitrile is added to 2-10% final concentration. The refolded protein is chromatographed on a Poros R1/H reversed phase column using a mobile buffer of 0.1% TFA with elution with a gradient of acetonitrile from 10 to 80%. Aliquots of fractions with A280 absorbance are analyzed on SDS polyacrylamide gels and fractions containing homogeneous refolded protein are pooled. Generally, the properly refolded species of most proteins are eluted at the lowest concentrations of acetonitrile since those species are the most compact with their hydrophobic interiors shielded from interaction with the reversed phase resin. Aggregated species are usually eluted at higher acetonitrile concentrations. In addition to resolving misfolded forms of proteins from the desired form, the reversed phase step also removes endotoxin from the samples.

Fractions containing the desired folded CHEPO polypeptide are pooled and the acetonitrile removed using a gentle stream of nitrogen directed at the solution. Proteins are formulated into 20 mM Hepes, pH 6.8 with 0.14 M sodium chloride and 4% mannitol by dialysis or by gel filtration using G25 Superfine (Pharmacia) resins equilibrated in the formulation buffer and sterile filtered.

EXAMPLE 4

Expression of CHEPO in mammalian cells

This example illustrates preparation of a potentially glycosylated form of CHEPO by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the CHEPO DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the CHEPO DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-CHEPO.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 g pRK5-CHEPO DNA is mixed with about 1 g DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 l of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 l of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 Ci/ml ^{35}S-cysteine and 200 Ci/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of CHEPO polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, CHEPO may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 g pRK5-CHEPO DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 g/ml bovine insulin and 0.1 g/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed CHEPO can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, CHEPO can be expressed in CHO cells. The pRK5-CHEPO can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of CHEPO polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed CHEPO can then be concentrated and purified by any selected method.

Epitope-tagged CHEPO may also be expressed in host CHO cells. The CHEPO may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged CHEPO insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged CHEPO can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

CHEPO may also be expressed in CHO and/or COS cells by a transient expression procedure or in CHO cells by another stable expression procedure.

Stable expression in CHO cells is performed using the following procedure. The proteins are expressed as an IgG construct (immunoadhesin), in which the coding sequences for the soluble forms (e.g. extracellular domains) of the respective proteins are fused to an IgG1 constant region sequence containing the hinge, CH2 and CH2 domains and/or is a poly-His tagged form.

Following PCR amplification, the respective DNAs are subcloned in a CHO expression vector using standard techniques as described in Ausubel et al., Current Protocols of Molecular Biology, Unit 3.16, John Wiley and Sons (1997). CHO expression vectors are constructed to have compatible restriction sites 5 and 3 of the DNA of interest to allow the convenient shuttling of cDNA s. The vector used expression in CHO cells is as described in Lucas et al., Nucl. Acids Res. 24:9 (1774-1779 (1996), and uses the SV40 early promoter/enhancer to drive expression of the cDNA of interest and dihydrofolate reductase (DHFR). DHFR expression permits selection for stable maintenance of the plasmid following transfection.

Twelve micrograms of the desired plasmid DNA is introduced into approximately 10 million CHO cells using commercially available transfection reagents Superfect (Quiagen), Dosper or Fugene (Boehringer Mannheim). The cells are grown as described in Lucas et al., supra. Approximately 3×10^{7} cells are frozen in an ampule for further growth and production as described below.

The ampules containing the plasmid DNA are thawed by placement into water bath and mixed by vortexing. The contents are pipetted into a centrifuge tube containing 10 mLs of media and centrifuged at 1000 rpm for 5 minutes. The supernatant is aspirated and the cells are resuspended in 10 mL of selective media (0.2 m filtered PS20 with 5% 0.2 m diafiltered fetal bovine serum). The cells are then aliquoted into a 100 mL spinner containing 90 mL of selective media. After 1-2 days, the cells are transferred into a 250 mL spinner filled with 150 mL selective growth medium and incubated at 37°C. After another 2-3 days, 250 mL, 500 mL and 2000 mL spinners are seeded with 3×10^{5} cells/mL. The cell media is exchanged with fresh media by centrifugation and resuspension in production medium. Although any suitable CHO media may be employed, a production medium described in U.S. Patent No. 5,122,469, issued June 16, 1992 may actually be used. A 3L production spinner is seeded at 1.2×10^{6} cells/mL. On day 0, the cell number pH ie determined. On day 1, the spinner is sampled and sparging with filtered air is commenced. On day 2, the spinner is sampled, the temperature shifted to 33°C, and 30 mL of 500 g/L glucose and 0.6 mL of 10% antifoam (e.g., 35% polydimethylsiloxane emulsion, Dow Corning 365 Medical Grade Emulsion) taken. Throughout the production, the pH is adjusted as necessary to keep it at

around 7.2. After 10 days, or until the viability dropped below 70%, the cell culture is harvested by centrifugation and filtering through a 0.22 m filter. The filtrate was either stored at 4°C or immediately loaded onto columns for purification.

For the poly-His tagged constructs, the proteins are purified using a Ni-NTA column (Qiagen). Before purification, imidazole is added to the conditioned media to a concentration of 5 mM. The conditioned media is pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer containing 0.3 M NaCl and 5 mM imidazole at a flow rate of 4-5 ml/min. at 4°C. After loading, the column is washed with additional equilibration buffer and the protein eluted with equilibration buffer containing 0.25 M imidazole. The highly purified protein is subsequently desalted into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80°C.

Immunoadhesin (Fc-containing) constructs are purified from the conditioned media as follows. The conditioned medium is pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column is washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein is immediately neutralized by collecting 1 ml fractions into tubes containing 275 L of 1 M Tris buffer, pH 9. The highly purified protein is subsequently desalted into storage buffer as described above for the poly-His tagged proteins. The homogeneity is assessed by SDS polyacrylamide gels and by N-terminal amino acid sequencing by Edman degradation.

EXAMPLE 5

Expression of CHEPO in Yeast

The following method describes recombinant expression of CHEPO in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of CHEPO from the ADH2/GAPDH promoter. DNA encoding CHEPO and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of CHEPO. For secretion, DNA encoding CHEPO can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, a native CHEPO signal peptide or other mammalian signal peptide, or, for example, a yeast alpha-factor or invertase secretory signal/leader sequence, and linker sequences (if needed) for expression of CHEPO.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant CHEPO can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing CHEPO may further be purified using selected column chromatography resins.

EXAMPLE 6

Expression of CHEPO in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of CHEPO in Baculovirus-infected insect cells.

The sequence coding for CHEPO is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the sequence encoding CHEPO or the desired portion of the coding sequence of CHEPO such as the sequence encoding the extracellular domain of a transmembrane protein or the sequence encoding the mature protein if the protein is extracellular is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 - 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression are performed as described by O'Reilley et al., Baculovirus expression vectors: A Laboratory Manual, Oxford: Oxford University Press (1994).

Expressed poly-his tagged CHEPO can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% glycerol, pH 7.8) and filtered through a 0.45 m filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or Western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged CHEPO are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) CHEPO can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7

Preparation of Antibodies that Bind CHEPO

This example illustrates preparation of monoclonal antibodies which can specifically bind CHEPO.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified CHEPO, fusion proteins containing CHEPO, and cells expressing recombinant CHEPO on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the CHEPO immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-CHEPO antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of CHEPO. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against CHEPO. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against CHEPO is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-CHEPO monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

EXAMPLE 8

Purification of CHEPO Polypeptides Using Specific Antibodies

Native or recombinant CHEPO polypeptides may be purified by a variety of standard techniques in the art of protein purification. For example, pro-CHEPO polypeptide, mature CHEPO polypeptide, or pre-CHEPO polypeptide is purified by immunoaffinity chromatography using antibodies specific for the CHEPO polypeptide of interest. In general, an immunoaffinity column is constructed by covalently coupling the anti-CHEPO polypeptide antibody to an activated chromatographic resin.

Polyclonal immunoglobulins are prepared from immune sera either by precipitation with ammonium sulfate or by purification on immobilized Protein A (Pharmacia LKB Biotechnology, Piscataway, N.J.). Likewise, monoclonal antibodies are prepared from mouse ascites fluid by ammonium sulfate precipitation or chromatography on immobilized Protein A. Partially purified immunoglobulin is covalently attached to a chromatographic resin such as CnBr-activated SEPHAROSE™ (Pharmacia LKB Biotechnology). The antibody is coupled to the resin, the resin is blocked, and the derivative resin is washed according to the manufacturer's instructions.

Such an immunoaffinity column is utilized in the purification of CHEPO polypeptide by preparing a fraction from cells containing CHEPO polypeptide in a soluble form. This preparation is derived by solubilization of the whole cell or of

a subcellular fraction obtained via differential centrifugation by the addition of detergent or by other methods well known in the art. Alternatively, soluble CHEPO polypeptide containing a signal sequence may be secreted in useful quantity into the medium in which the cells are grown.

A soluble CHEPO polypeptide-containing preparation is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CHEPO polypeptide (*e.g.*, high ionic strength buffers in the presence of detergent). Then, the column is eluted under conditions that disrupt antibody/CHEPO polypeptide binding (*e.g.*, a low pH buffer such as approximately pH 2-3, or a high concentration of a chaotrope such as urea or thiocyanate ion), and CHEPO polypeptide is collected.

EXAMPLE 9

Drug Screening

This invention is particularly useful for screening compounds by using CHEPO polypeptides or binding fragment thereof in any of a variety of drug screening techniques. The CHEPO polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the CHEPO polypeptide or fragment. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between CHEPO polypeptide or a fragment and the agent being tested. Alternatively, one can examine the diminution in complex formation between the CHEPO polypeptide and its target cell or target receptors caused by the agent being tested.

Thus, the present invention provides methods of screening for drugs or any other agents which can affect a CHEPO polypeptide-associated disease or disorder. These methods comprise contacting such an agent with an CHEPO polypeptide or fragment thereof and assaying (I) for the presence of a complex between the agent and the CHEPO polypeptide or fragment, or (ii) for the presence of a complex between the CHEPO polypeptide or fragment and the cell, by methods well known in the art. In such competitive binding assays, the CHEPO polypeptide or fragment is typically labeled. After suitable incubation, free CHEPO polypeptide or fragment is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability of the particular agent to bind to CHEPO polypeptide or to interfere with the CHEPO polypeptide/cell complex.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to a polypeptide and is described in detail in WO 84/03564, published on September 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. As applied to a CHEPO polypeptide, the peptide test compounds are reacted with CHEPO polypeptide and washed. Bound CHEPO polypeptide is detected by methods well known in the art. Purified CHEPO polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies can be used to capture the peptide and immobilize it on the solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding CHEPO polypeptide specifically compete with a test compound for binding to CHEPO polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CHEPO polypeptide.

EXAMPLE 10

Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptide of interest (*i.e.*, a CHEPO polypeptide) or of small molecules with which they interact, *e.g.*, agonists, antagonists, or inhibitors. Any of these examples can be used to fashion drugs which are more active or stable forms of the CHEPO polypeptide or which enhance or interfere with the function of the CHEPO polypeptide *in vivo* (*c.f.*, Hodgson, Bio/Technology, 9: 19-21 (1991)).

In one approach, the three-dimensional structure of the CHEPO polypeptide, or of an CHEPO polypeptide-inhibitor complex, is determined by x-ray crystallography, by computer modeling or, most typically, by a combination of the two approaches. Both the shape and charges of the CHEPO polypeptide must be ascertained to elucidate the structure and to determine active site(s) of the molecule. Less often, useful information regarding the structure of the CHEPO polypeptide may be gained by modeling based on the structure of homologous proteins. In both cases, relevant structural information is used to design analogous CHEPO polypeptide-like molecules or to identify efficient inhibitors. Useful examples of rational drug design may include molecules which have improved activity or stability as shown by Braxton and Wells, Biochemistry, 31:7796-7801 (1992) or which act as inhibitors, agonists, or antagonists of native peptides as shown by Athauda *et al.*, J. Biochem., 113:742-746 (1993).

It is also possible to isolate a target-specific antibody, selected by functional assay, as described above, and then to solve its crystal structure. This approach, in principle, yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced peptides. The isolated peptides would then act as the pharmacore.

By virtue of the present invention, sufficient amounts of the CHEPO polypeptide may be made available to perform such analytical studies as X-ray crystallography. In addition, knowledge of the CHEPO polypeptide amino acid sequence provided herein will provide guidance to those employing computer modeling techniques in place of or in addition to x-ray crystallography.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention as claimed. Various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule having at least about 80% nucleic acid sequence identity to (a) a DNA molecule encoding a CHEPO polypeptide comprising the sequence of amino acid residues from about 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2), or (b) the complement of the DNA molecule of (a).

2. The isolated nucleic acid molecule of Claim 1 comprising nucleotides 1 or about 82 to about 579 of Figure 2 (SEQ ID NO:3).

3. The isolated nucleic acid molecule of Claim 1 comprising the nucleotide sequence of Figure 2 (SEQ ID NO:3).

4. The isolated nucleic acid molecule of Claim 1 comprising a nucleotide sequence that encodes the sequence of amino acid residues from about 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2).

5. An isolated nucleic acid molecule encoding a CHEPO polypeptide comprising DNA that hybridizes to the complement of the nucleic acid sequence that encodes amino acids 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2).

6. The isolated nucleic acid molecule of Claim 5, wherein the nucleic acid that encodes amino acids 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2) comprises nucleotides 1 or about 82 to about 579 of Figure 2 (SEQ ID NO:3).

7. The isolated nucleic acid molecule of Claim 5, wherein the hybridization occurs under stringent hybridization and wash conditions.

8. An isolated nucleic acid molecule comprising (a) DNA encoding a polypeptide scoring at least 80% positives when compared to the sequence of amino acid residues of from 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2), or (b) the complement of the DNA of (a).

9. A vector comprising the nucleic acid molecule of Claim 1.

10. The vector of Claim 9, wherein said nucleic acid molecule is operably linked to control sequences recognized by a host cell transformed with the vector.

11. A host cell comprising the vector of Claim 9.

12. The host cell of Claim 11, wherein said cell is a CHO cell.

13. The host cell of Claim 11, wherein said cell is an *E. coli*.

14. The host cell of Claim 11, wherein said cell is a yeast cell.

15. A process for producing a CHEPO polypeptide comprising culturing the host cell of Claim 11 under conditions suitable for expression of said CHEPO polypeptide and recovering said CHEPO polypeptide from the cell culture.

16. An isolated CHEPO polypeptide comprising an amino acid sequence comprising at least about 80% sequence identity to the sequence of amino acid residues from about 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2).

17. The isolated CHEPO polypeptide of Claim 16 comprising amino acid residues 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2).

18. An isolated CHEPO polypeptide scoring at least 80% positives when compared to the sequence of amino acid residues from 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2).

19. An isolated CHEPO polypeptide comprising the sequence of amino acid residues from 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2), or a fragment thereof sufficient to provide a binding site for an anti-CHEPO antibody.

20. An isolated polypeptide produced by (i) hybridizing a test DNA molecule under stringent conditions with (a) a DNA molecule encoding a CHEPO polypeptide comprising the sequence of amino acid residues from 1 or about 28 to about 193 of Figure 2 (SEQ ID NO:2), or (b) the complement of the DNA molecule of (a), (ii) culturing a host cell comprising said test DNA molecule under conditions suitable for the expression of said polypeptide, and (iii) recovering said polypeptide from the cell culture.

21. The isolated polypeptide of Claim 20, wherein said test DNA has at least about 80% sequence identity to (a) or (b).

22. A chimeric molecule comprising a CHEPO polypeptide fused to a heterologous amino acid sequence.

23. The chimeric molecule of Claim 22, wherein said heterologous amino acid sequence is an epitope tag sequence.

24. The chimeric molecule of Claim 22, wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

25. An antibody which specifically binds to a CHEPO polypeptide.

26. The antibody of Claim 24, wherein said antibody is a monoclonal antibody.

27. The antibody of Claim 24, wherein said antibody is a humanized antibody.

28. An agonist to a CHEPO polypeptide.

29. An antagonist to a CHEPO polypeptide.

30. A composition of matter comprising (a) a CHEPO polypeptide, (b) an agonist to a CHEPO polypeptide, (c) an antagonist to a CHEPO polypeptide, or (d) an anti-CHEPO antibody in admixture with a pharmaceutically acceptable carrier.

31. A method of inducing erythropoiesis in a mammal, said method comprising administering to said mammal an effective amount of a CHEPO polypeptide or an agonist thereto, wherein erythropoiesis in said mammal is induced.

32. A method of inhibiting erythropoiesis in a mammal, said method comprising administering to said mammal an effective amount of an antagonist to a CHEPO polypeptide, wherein erythropoiesis in said mammal is inhibited.

33. The method according to Claim 31, wherein said antagonist is an anti-CHEPO antibody.

34. A CHEPO polypeptide comprising an amino acid sequence selected from the group consisting of:

(1) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNXSXQQAVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:18);

(2) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNXSXQQAVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:19);

(3) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNXTXQQAVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:20);

(4) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNXTXQQAVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:21);

(5) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMNXSXQAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:22);

(6) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMNXSXQAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:23);

(7) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMNXTXQAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:24);

(8) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMNXTXQAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:25);

(9) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMENXSXAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:26);

(10) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMENXSXAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:27);

(11) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMENXTXAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:28);

(12) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMENXTXAVEVWQGLALLSEAVLRGQALL VNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACR TGDR (SEQ ID NO:29);

(13) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMEVNXSXVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:30);

(14) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMEVNXSXVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:31);

(15) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMEVNXTXVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:32); and

(16) APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMEVNXTXVEVWQGLALLSEAVLRGQALLV NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRT GDR (SEQ ID NO:33), wherein X is any amino acid except for proline.

35. The CHEPO polypeptide according to Claim 34, which comprises an amino acid sequence selected from the group consisting of:

(1) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNX SXQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:34);

(2) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNX SXQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:35);

(3) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNX TXQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:36);

(4) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRNX TXQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:37);

(5) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMN XSXQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:38);

(6) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMN XSXQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:39);

(7) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMN XTXQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:40);

(8) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMN XTXQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:41);

(9) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME NXSXAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:42);

(10) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME NXSXAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:43);

(11) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME NXTXAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:44);

(12) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME NXTXAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:45);

(13) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME VNXSXVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:46);

(14) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME VNXSXVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:47);

(15) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME VNXTXVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:48); and

(16) MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRME VNXTXVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAKKEAISPPDAASAAPLRTITAD TFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO:49), wherein X is any amino acid except for proline.

36. A chimeric molecule comprising a CHEPO polypeptide of Claim 34 or 35 fused to a heterologous amino acid sequence.

37. The chimeric molecule of Claim 36, wherein said heterologous amino acid sequence is an epitope tag sequence.

38. The chimeric molecule of Claim 36, wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

FIGURE 1A

```
   1 CCCCCTGGAC AGCCGCCCTC TCCTCCAGGC CCGTGGGGCT GGCCCTGCAC CGCCGAGCTT CCCGGGATGA GGGCCCCCGG TGTGGTCACC CGGCGCGCCC
     GGGGGACCTG TCGGCGGGAG AGGAGGTCCG GGCACCCCGA CCGGGACGTG GCGGCTCGAA GGGCCCTACT CCCGGGGGCC ACACCAGTGG GCCGCGCGGG

 101 CAGGTCGCTG AGGGACCCCG GCCAGGCGCG GAGATGGGGG TGCACGGTGA GTACTCGCGG GCTGGGCGCT CCCGCCCGCC CGGGTCCCTG TTTGAGCGGG
     GTCCAGCGAC TCCCTGGGGC CGGTCCGCGC CTCTACCCCC ACGTGCCACT CATGAGCGCC CGACCCGCGA GGGCGGGCGG GCCCAGGGAC AAACTCGCCC
   1                                   M  G  V  H
                                       ^Exon 1
                                       ^Met

 201 GATTTAGCGC CCGGGCTATT GGCCGGGAGG TGGCTGGGTT CAAGGACCGG CGACTTGTCA AGGACCCCGG AAGGGGGAGG GGGGTGGGGC AGCCTCCACG
     CTAAATCGCG GGCCCGATAA CCGGCCCTCC ACCGACCCAA GTTCCTGGCC GCTGAACAGT TCCTGGGGCC TTCCCCCTCC CCCCACCCCG TCGGAGGTGC

 301 TGCCAGCGGG GACTTGGGGG AGTCCTTGGG GATGGCAAAA ACCTGACCTG TGAAGGGGAC ACAGTTTGGG GGTTGAGGGG AAGAAGGTTT GGGGGTTCTG
     ACGGTCGCCC CTGAACCCCC TCAGGAACCC CTACCGTTTT TGGACTGGAC ACTTCCCCTG TGTCAAACCC CCAACTCCCC TTCTTCCAAA CCCCCAAGAC

 401 CTGTGCCAST GGAGAGGAAG CTGATAAGCT GATAACCTGG GCGCTGGAGC CACCACTTAT CTGCCAGAGG GNNNNTGGTA GCTGGGGGTG GGGTGTGCAC
     GACACGGTSA CCTCTCCTTC GACTATTCGA CTATTGGACC CGCGACCTCG GTGGTGAATA GACGGTCTCC CNNNNACCAT CGACCCCCAC CCCACACGTG

 501 ACGGCAGCAG GATTGAATGA AGGCCAGGGA GGCAGCACCT GAGTGCTTGC ATGGTTGGGG ACAGGAAGGA CGAGCTGGGG CAGAGACGTG GGGATGAAGG
     TGCCGTCGTC CTAACTTACT TCCGGTCCCT CCGTCGTGGA CTCACGAACG TACCAACCCC TGTCCTTCCT GCTCGACCCC GTCTCTGCAC CCCTACTTCC

 601 AAGCTGTCCT TCCACAGCCA CCCTTCTCCC TCCCCGCCTG ACTCTCAGCC TGGCTATCTC TTCTAGAATG TCCTGCCTGG CTGTGGCTTC TCCTGTCCCT
     TTCGACAGGA AGGTGTCGGT GGGAAGAGGG AGGGGCGGAC TGAGAGTCGG ACCGATAGAG AAGATCTTAC AGGACGGACC GACACCGAAG AGGACAGGGA
   1                                                                   E  C  P  A  W  L  W  L  L  L  S  L
                                                                       ^exon 2

 701 GCTGTCGCTC CCTCTGGGCC TCCCAGTCCT GGGCGCCCCA CCACGCCTCA TCTGTGACAG CCGAGTCCTG GAGAGGTACC TCTTGGAGGC CAAGGAGGCC
     CGACAGCGAG GGAGACCCGG AGGGTCAGGA CCCGCGGGGT GGTGCGGAGT AGACACTGTC GGCTCAGGAC CTCTCCATGG AGAACCTCCG GTTCCTCCGG
  13 L  S  L  P  L  G  L  P  V  L  G  A  P  P  R  L  I  C  D  S  R  V  L  E  R  Y  L  L  E  A  K  E  A

 801 GAGAATATCA CGGTGAGACC CCTTCCCCAG CACATTCCAC AGAACTCACG CTCAGGGCTT CAGGGAACTC CTCCCAGATC CAGGAACCTG GCACTTGGTT
     CTCTTATAGT GCCACTCTGG GGAAGGGGTC GTGTAAGGTG TCTTGAGTGC GAGTCCCGAA GTCCCTTGAG GAGGGTCTAG GTCCTTGGAC CGTGAACCAA
  46 E  N  I  T
```


SUBSTITUTE SHEET (RULE 26)

FIGURE 1B

```
 901 TGGGGTGGAG TTGGGAAGCT AGACACTGCC CCCCTACATA AGAATAAGTC TGGTGGCCCC AAACCATACC TGGAAACTAG GCAAGGAGCA AAGCCAGCAG
     ACCCCACCTC AACCCTTCGA TCTGTGACGG GGGGATGTAT TCTTATTCAG ACCACCGGGG TTTGGTATGG ACCTTTGATC CGTTCCTCGT TTCGGTCGTC

1001 ATCCTACGGC CTGTGGGCCA GGGCCAGAGC CTTCAGGGAC CCTTGACTCC CCGGGCTGTG TGCATTTCAG ACGGGCTGTG CCGAACACTG CAGCTTGAAT
     TAGGATGCCG GACACCCGGT CCCGGTCTCG GAAGTCCCTG GGAACTGAGG GGCCCGACAC ACGTAAAGTC TGCCCGACAC GGCTTGTGAC GTCGAACTTA
                                                                    T  G  C  A  E  H  C  S  L  N
                                                                                ^exon 3
1101 GAGAATATCA CTGTCCCAGA CACCAAAGTT AATTTCTATG CCTGGAAGAG GATGGAGGTG AGTTCCTTTT TTTTTTTTTT TCCTTTCTTT TGGAGAATCT
     CTCTTATAGT GACAGGGTCT GTGGTTTCAA TTAAAGATAC GGACCTTCTC CTACCTCCAC TCAAGAAAA AAAAAAAAAA AGGAAAGAAA ACCTCTTAGA
      11 E  N  I  T  V  P  D  T  K  V  N  F  Y  A  W  K  R  M  E

1201 CATTTGCGAG CCTGATTTTG GATGAAAGGG AGAATGATCG AGGGAAAGGT AAAATGGAGC AGCAGAGATG AGGCTGCCTG GGCGCAGAGG CTCACGTCTA
     GTAAACGCTC GGACTAAAAC CTACTTTCCC TCTTACTAGC TCCCTTTCCA TTTTACCTCG TCGTCTCTAC TCCGACGGAC CCGCGTCTCC GAGTGCAGAT

1301 TAATCCCAGG CTGAGATGGC CGAGATGGGA GAATTGCTTG AGCCCTGGAG TTTCAGACCA ACCTGGGCAG CATAGTGAGA TCCCCCATCT CTACAAACAT
     ATTAGGGTCC GACTCTACCG GCTCTACCCT CTTAACGAAC TCGGGACCTC AAAGTCTGGT TGGACCCGTC GTATCACTCT AGGGGGTAGA GATGTTTGTA

1401 TTAAAAAAAT TAGTCAGGTG AGGTGGTGCA TGGTGGTAGT CCCAGATATT TGGAAGGCTG AGGCGGGAGG ATCGCTTGAG CCCAGGAATT TGAGGCTGCA
     AATTTTTTTA ATCAGTCCAC TCCACCACGT ACCACCATCA GGGTCTATAA ACCTTCCGAC TCCGCCCTCC TAGCGAACTC GGGTCCTTAA ACTCCGACGT

1501 GTGAGCTGTG ATCACACCAC TGCACTCCAG CCTCAGTGAC AGAGTGAGGC CCTGTCTCAA AAAGAAAAG AAAAAAGAAA AATAATGAGG GCTGTATGGA
     CACTCGACAC TAGTGTGGTG ACGTGAGGTC GGAGTCACTG TCTCACTCCG GGACAGAGTT TTTTCTTTTC TTTTTTCTTT TTATTACTCC CGACATACCT

1601 ATAATTCAT TATTCATTCA CTCACTCATT CATTCATTCA TTCATTCNNN NNNTCTTATT GCATACCTCT GTTTGCTCAG CTTGGTGCTT GGGGCTGCTG
     TATGTAAGTA ATAAGTAAGT GAGTGAGTAA GTAAGTAAGT AAGTAAGNNN NNNAGAATAA CGTATGGAGA CAACGAGTCC GAACCACGAA CCCCGACGAC

1701 AGGGGCAGGA GGGAGAGGGT GGCATGGGTC AGCTGACTCC CAGAGTCCAC TCCCTGTAGG TCAGGCAGCA GGCCGTAGAA GTCTGGCAGG GCCTGGCCCT
     TCCCCGTCCT CCCTCTCCCA CCGTACCCAG TCGACTGAGG GTCTCAGGTG AGGGACATCC AGTCCGTCGT CCGGCATCTT CAGACCGTCC CGGACCGGGA
                                                                    V  R  Q  Q  A  V  E  V  W  Q  G  L  A  L
                                                                                ^exon 4
1801 GCCCTCGGAA GCTGTCCTGC GGGGCCAGGC CCTGTTGGTC AACTCTTCCC AGCCGTGGGA GCCCCTGCAG CTGCATGTGG ATAAAGCCGT CAGTGGCCTT
     CGAGAGCCTT CGACAGGACG CCCCGGTCCG GGACAACCAG TTGAGAAGGG TCGGCACCCT CGGGGACGTC GACGTACACC TATTTCGGCA GTCACCGGAA
      25 L  S  E  A  V  L  R  G  Q  A  L  L  V  N  S  S  Q  P  W  E  P  L  Q  L  H  V  D  K  A  V  S  G  L

1901 CGCAGCCTCA CCACTCTGCT TCGGGCTCTG GGAGCCCAGG TGAGTAGGAG CGGACACTTC TGCTTGCCCT TTCTGTAAGA AAGGGAGAAG GGTCTTGCTA
     GCGTCGGAGT GGTGAGACGA AGCCCGAGAC CCTCGGGTCC ACTCATCCTC GCCTGTGAAG ACGAACGGGA AAGACATTCT TTCCCTCTTC CCAGAACGAT
      48 R  S  L  T  T  L  L  R  A  L  G  A  Q
```

FIGURE 1C

```
2001 AGGAGTACAG GAACTGTCCG TATTCCTTCC CCTTCTGTGG CACTGCAGCG ACCTCCTGTT TTCTCCTTGG CAGAAGGAAG CCATCTCCCC TCCAGATGCG
     TCCTCATGTC CTTGACAGGC ATAAGGAAGG GGAAGACACC GTGACGTCGC TGGAGGACAA AAGAGGAACC GTCTTCCTTC GGTAGAGGGG AGGTCTACGC
       1                                                                    K  E  A  I  S  P  P  D  A
                                                                          ^exon 5

2101 GCCTCAGCTG CTCCACTCCG AACAATCACT GCTGACACTT TCCGCAAACT CTTCCGAGTC TACTCCAATT TCCTCCGGGG AAAGCTGAAG CTGTACACAG
     CGGAGTCGAC GAGGTGAGGC TTGTTAGTGA CGACTGTGAA AGGCGTTTGA GAAGGCTCAG ATGAGGTTAA AGGAGGCCCC TTTCGACTTC GACATGTGTC
      10 A  S  A  A  P  L  R  T  I  T  A  D  T  F  R  K  L  F  R  V  Y  Y  S  N  F  L  R  G  K  L  K  L  Y  T  G

2201 GGGAGGCCTG CAGGACAGGG GACAGATGAC CAGGTGTGTC CACCTGGGCA TATCCACCAC CTCCCTCACC AACATTGCTT GTGCCACACC CTCCCCCGCC
     CCCTCCGGAC GTCCTGTCCC CTGTCTACTG GTCCACACAG GTGGACCCGT ATAGGTGGTG GAGGGAGTGG TTGTAACGAA CACGGTGTGG GAGGGGGCGG
       44 E  A  C  R  T  G  D  R  O

2301 ACTCCTGAAC CCCGTCGAGG AGCTCTCAG
     TGAGGACTTG GGGCAGCTCC TCGAGAGTC
```

FIGURE 2

```
  1 ATGGGGGTGC ACGAATGTCC TGCCTGGCTG TGGCTTCTCC TGTCCCTGCT GTCGCTCCCT CTGGGCCTCC CAGTCCTGGG CGCCCCACCA CGCCTCATCT
    TACCCCCACG TGCTTACAGG ACGGACCGAC ACCGAAGAGG ACAGGGACGA CAGCGAGGGA GACCCGGAGG GTCAGGACCC GCGGGGTGGT GCGGAGTAGA
  1 M  G  V  H     E  C  P     A  W  L     W  L  L  L     S  L  L     S  L  P     L  G  L  P     V  L  G     A  P  P     R  L  I  C

101 GTGACAGCCG AGTCCTGGAG AGGTACCTCT TGGAGGCCAA GGAGGCCGAG AATATCACGA CGGGCTGTGC CGAACACTGC AGCTTGAATG AGAATATCAC
    CACTGTCGGC TCAGGACCTC TCCATGGAGA ACCTCCGGTT CCTCCGGCTC TTATAGTGCT GCCCGACACG GCTTGTGACG TCGAACTTAC TCTTATAGTG
 35   D  S  R     V  L  E     R  Y  L  L     E  A  K     E  A  E     N  I  T  T     G  C  A     E  H  C     S  L  N  E     N  I  T
                                                                      ^*                                                   ^*

201 TGTCCCAGAC ACCAAAGTTA ATTTCTATGC CTGGAAGAGG ATGGAGGTCA GGCAGCAGGC CGTAGAAGTC TGGCAGGGCC TGGCCCTGCT CTCGGAAGCT
    ACAGGGTCTG TGGTTTCAAT TAAAGATACG GACCTTCTCC TACCTCCAGT CCGTCGTCCG GCATCTTCAG ACCGTCCCGG ACCGGGACGA GAGCCTTCGA
 68 V  P  D     T  K  V  N     F  Y  A     W  K  R     M  E  V  R     Q  Q  A     V  E  V     W  Q  G  L     A  L  L     S  E  A

301 GTCCTGCGGG GCCAGGCCCT GTTGGTCAAC TCTTCCCAGC CGTGGGAGCC CCTGCAGCTG CATGTGGATA AAGCCGTCAG TGGCCTTCGC AGCCTCACCA
    CAGGACGCCC CGGTCCGGGA CAACCAGTTG AGAAGGGTCG GCACCCTCGG GGACGTCGAC GTACACCTAT TTCGGCAGTC ACCGGAAGCG TCGGAGTGGT
101 V  L  R  G     Q  A  L     L  V  N     S  S  Q  P     W  E  P     L  Q  L     H  V  D  K     A  V  S     G  L  R     S  L  T  T
                                  ^*

401 CTCTGCTTCG GGCTCTGGGA GCCCAGAAGG AAGCCATCTC CCCTCCAGAT GCGGCCTCAG CTGCTCCACT CCGAACAATC ACTGCTGACA CTTTCCGCAA
    GAGACGAAGC CCGAGACCCT CGGGTCTTCC TTCGGTAGAG GGGAGGTCTA CGCCGGAGTC GACGAGGTGA GGCTTGTTAG TGACGACTGT GAAAGGCGTT
135   L  L  R     A  L  G     A  Q  K  E     A  I  S     P  P  D     A  A  S  A     A  P  L     R  T  I     T  A  D  T     F  R  K
                                ^Polymorphism in CRL 1609 (A or C). A changes the protein sequence to K.

501 ACTCTTCCGA GTCTACTCCA ATTTCCTCCG GGGAAAGCTG AAGCTGTACA CAGGGGAGGC CTGCAGGACA GGGGACAGAT GACCA
    TGAGAAGGCT CAGATGAGGT TAAAGGAGGC CCCTTTCGAC TTCGACATGT GTCCCCTCCG GACGTCCTGT CCCCTGTCTA CTGGT
168 L  F  R     V  Y  S  N     F  L  R     G  K  L     K  L  Y  T     G  E  A     C  R  T     G  D  R  O
```

FIGURE 3

```
human   MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE
        **************************************************
chepo   MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE

human   NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA
        ********************************** ***************
chepo   NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVRQQAVEVWQGLALLSEA

human   VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD
        ****************************************** ********
chepo   VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAXKEAISPPD

human   AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
        *******************************************
chepo   AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
```

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/US 00/12370

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/11 C12N15/12 C12N15/62 C07K14/505 C07K16/22
G01N33/50 G01N33/53 A61K48/00 A61K38/18 A61P7/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N A61K A61P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, MEDLINE, STRAND, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 24160 A (BRIGHAM & WOMENS HOSPITAL) 27 October 1994 (1994-10-27) Note: 89.9 (89.4) % aa seq identity of monkey EPO with SEQ ID NO:2 (5) in 189 aa overlap, 86.9 % nt seq identity of cat EPO cDNA with SEQ ID NO:3 in 573 aa overlap. the whole document page 7, line 8 -page 8, line 12 example 1 page 55, table IV figure 6 --- -/--	1,8-15, 20,21

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 August 2000	15/09/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	van de Kamp, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/US 00/12370

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WEN D ET AL.: "Erythropoietin structure-function relationships: high degree of sequence homology among mammals" BLOOD, vol. 82, no. 5, 1 September 1993 (1993-09-01), pages 1507-1516, XP000938929 Note: 90.2 (89.6) % aa seq identity of Macaca mulatta EPO with SEQ ID NO:2 (5) in 193 aa overlap. the whole document	1,8,20, 21
X	WO 95 05465 A (AMGEN INC) 23 February 1995 (1995-02-23) Note: 100 % aa seq identity of Asn57Thr59 EPO with SEQ ID NO:32. page 5, line 10-25 page 41, table 3, analog N9 page 56, table 6, Asn57Thr59 claims 1-8,10	1,8-15, 20,21, 34-38
X	EP 0 410 246 A (BEHRINGWERKE AG) 30 January 1991 (1991-01-30) the whole document	28-33
X	WO 95 13376 A (AMGEN INC ;UNIV SOUTHERN CALIFORNIA (US)) 18 May 1995 (1995-05-18) Note: 99.5 % aa seq identity of human EPO with SEQ ID NO:2 in 193 aa overlap, 99.5 % nt seq identity of human EPO cDNA with SEQ ID NO:3 in 585 nt overlap. page 38-40	1,8-15, 20,21
X	WO 85 02610 A (KIRIN AMGEN INC) 20 June 1985 (1985-06-20) Note: 91.2 (90.7) % aa seq identity of Cynomolgus monkey EPO with SEQ ID NO:2 (5) in 193 aa overlap. examples 3,10,12	1,8-15, 20,21
A	WO 99 02709 A (STROM TERRY B ;ZHENG XIN XIAO (US); BETH ISRAEL HOSPITAL (US); SYT) 21 January 1999 (1999-01-21) page 22, line 24 -page 29, line 7	22,24, 36,38
P,X	WO 00 24893 A (AMGEN INC) 4 May 2000 (2000-05-04) Note: 100 % aa seq identity of Asn55Thr57 EPO with SEQ ID NO:24. page 7, line 2-10 page 27, table 1, analog N54 claims 9,11,23-27	1,8-15, 20,21, 34-38

3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 28, 30(b), 31(b)

Remark (2): Present claims 28-32 relate to an extremely large number of possible CHEPO agonists and antagonists. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the antagonists claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to anti-CHEPO antibodies, as claimed in claims 30 and 33. Claim 28, and those parts of claims 30 and 31 where reference is made to agonists to a CHEPO polypeptide, have not been searched.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/US 00/12370

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9424160 A	27-10-1994	AU 6709794 A	08-11-1994
WO 9505465 A	23-02-1995	AT 155796 T	15-08-1997
		AU 677097 B	10-04-1997
		AU 7632794 A	14-03-1995
		CN 1105030 A	12-07-1995
		CZ 9500917 A	13-03-1996
		DE 69404401 D	04-09-1997
		DE 69404401 T	19-02-1998
		DK 640619 T	02-02-1998
		EP 0640619 A	01-03-1995
		ES 2105442 T	16-10-1997
		FI 951792 A	18-05-1995
		GR 3024815 T	30-01-1998
		HK 1001589 A	26-06-1998
		HU 72849 A	28-05-1996
		JP 11155584 A	15-06-1999
		JP 2938572 B	23-08-1999
		JP 8506023 T	02-07-1996
		LV 10972 A	20-12-1995
		LV 10972 B	20-10-1996
		NO 951445 A	16-06-1995
		NZ 273134 A	24-11-1997
		SI 640619 T	28-02-1998
		SK 50295 A	09-08-1995
		ZA 9406122 A	20-03-1995
EP 0410246 A	30-01-1991	DE 3924746 A	31-01-1991
		AT 161577 T	15-01-1998
		AU 648506 B	28-04-1994
		AU 5977890 A	31-01-1991
		CA 2021912 A	27-01-1991
		DE 59010787 D	05-02-1998
		ES 2111524 T	16-03-1998
		GR 3026096 T	29-05-1998
		JP 3081294 A	05-04-1991
		KR 175657 B	01-04-1999
		PT 94813 A,B	20-03-1991
		US 5712370 A	27-01-1998
		US 5106954 A	21-04-1992
WO 9513376 A	18-05-1995	AU 1095595 A	29-05-1995
		CA 2183551 A	18-05-1995
		EP 0737252 A	16-10-1996
WO 8502610 A	20-06-1985	US 4703008 A	27-10-1987
		AT 54940 T	15-08-1990
		AU 1007495 A	06-04-1995
		AU 657555 B	16-03-1995
		AU 2042192 A	08-10-1992
		AU 3612597 A	18-12-1997
		AU 600650 B	23-08-1990
		AU 3746785 A	26-06-1985
		AU 5272293 A	24-03-1994
		AU 5750490 A	04-10-1990
		AU 6981798 A	20-08-1998
		CA 1339047 A	27-05-1997
		CY 1643 A	14-05-1993

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/12370

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 8502610 A		DE 3482828 D	30-08-1990
		DK 365185 A	11-10-1985
		EP 0148605 A	17-07-1985
		ES 538519 D	01-05-1988
		ES 8802329 A	16-07-1988
		HK 20093 A	19-03-1993
		IL 73785 A	15-11-1992
		IL 96581 A	13-05-1993
		IL 96582 A	04-04-1993
		IL 100935 A	04-04-1993
		JP 2708099 B	04-02-1998
		JP 8269096 A	15-10-1996
		JP 6093000 A	05-04-1994
		JP 7076239 B	16-08-1995
		JP 2957974 B	06-10-1999
		JP 10095799 A	14-04-1998
		JP 3017962 B	13-03-2000
		JP 10072366 A	17-03-1998
		JP 1055190 A	02-03-1989
		JP 2107224 C	06-11-1996
		JP 6055136 B	27-07-1994
		JP 1975769 C	27-09-1995
		JP 3198792 A	29-08-1991
		JP 4035159 B	10-06-1992
		JP 2655750 B	24-09-1997
		JP 3259098 A	19-11-1991
		JP 11253188 A	21-09-1999
		JP 2017156 B	19-04-1990
		JP 61501627 T	07-08-1986
		MX 9203598 A	01-09-1992
		NZ 210501 A	27-08-1991
		NZ 227326 A	27-08-1991
		NZ 227327 A	27-08-1991
		SG 92891 G	14-02-1992
		US 5955422 A	21-09-1999
		US 5441868 A	15-08-1995
		US 5756349 A	26-05-1998
WO 9902709 A	21-01-1999	AU 8182298 A	08-02-1999
WO 0024893 A	04-05-2000	AU 1124100 A	15-05-2000